

Python (Semester 2 2024)

## **Astronomical Data Sources**

**Nina Hernitschek**

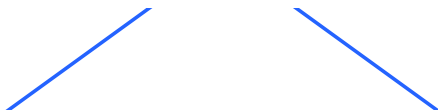
Centro de Astronomía CITEVA  
Universidad de Antofagasta

October 14, 2024

# Motivation

One of the cornerstones of research is **reproducibility**.

Reproducibility strongly depends on the availability of data.



surveys

data services

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# Motivation

We will see an introduction to **accessing astronomical data with Python**.

In addition, today's session we can also use for answering questions regarding the project (if any).

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# Surveys as Astronomical Data Sources

Motivation

Surveys

Services

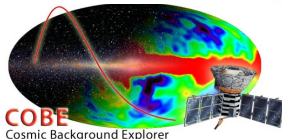
Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

Astronomical surveys often share their full data sets. In some cases, these are shared via the observatory that acquired them, for example, the all-sky data acquired with Planck, Wilkinson Microwave Anisotropy Probe, and COBE.

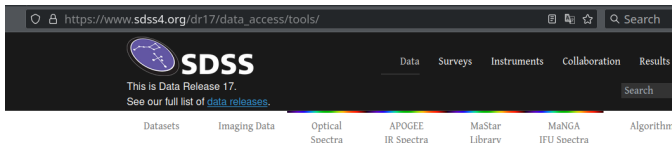
Other surveys serve their own data. Examples include the SDSS, 2MASS, UKIDSS.



# Surveys as Astronomical Data Sources

Example: SDSS

<https://www.sdss4.org>



## Available Tools for Data Access

### Table of Contents

Available Tools for Data Access

Catalog Archive Server (CAS)

Science Archive Server (SAS)

Direct file downloads

This page describes several tools for accessing the SDSS data: the Catalog Archive Server (CAS), the Science Archive Server (SAS), and direct data file access via rsync, wget, or http.

### Catalog Archive Server (CAS)

Catalog data are available through [SkyServer](#) (also known as the Catalog Archive Server, or CAS). SkyServer also offers interactive tools to browse through SDSS images with links to associated spectra and catalog data about the objects on the images.

<a href="#">DR17 SkyServer</a>	Interactive query tools for CAS
<a href="#">SkyServer Imaging Query Form</a>	Simple image query form interface
<a href="#">SkyServer Spectroscopic Query Form</a>	Simple spectroscopic query form interface
<a href="#">SkyServer Cross-ID</a>	Match list of RA/Dec or IDs to imaging or spectroscopic catalog

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# Surveys as Astronomical Data Sources

Example: SDSS

<https://skyserver.sdss.org>

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

The screenshot shows the SDSS SkyServer Search Tools interface. The browser address bar displays <https://skyserver.sdss.org/dr18/SearchTools/sql>. The page has a dark blue header with navigation links: Home, Visual Tools, Search Tools, CrossMatch Tools, More Tools, Support, and CasJobs. A left sidebar lists search methods: Overview, SQL Search, Rectangular Search, Radial Search, Imaging Search, Spectroscopic Search, and Spectroscopic Search. The main content area is titled "SQL Search" and contains a text editor with a SQL query. The query is a JOIN between the 'PhotoObj' and 'SpecObj' tables, selecting columns like p.objid, p.ra, p.dec, p.p, p.g, p.r, p.i, p.z, p.run, p.runRef, p.conc, p.field, s.specobjid, s.class, s.z, s.redshift, s.plate, s.hjd, s.fiberid, and s.p. The query filters for objects with p.ra between 0 and 15.6 and p.dec between 0 and 20. Below the query editor, there are buttons for "Output Format" and a list of supported formats: HTML, CSV, XML, JSON, OVD, Table, and FITS. On the right side, there is a "Sample Queries" section with a list of query categories: Basic SQL, SQL Jujitsu, Black Hole Mapper, Milky Way Mapper, Miscellaneous, Variability Queries, General Astronomy, Galaxies, Stars, Quasars, BOSS, and APOGEE.

SQL Search

```
1 -- This query does a table JOIN between the Imaging (PhotoObj) and spectra
2 -- (SpecObj) tables and includes the necessary columns in the SELECT to upload
3 -- the results to the SAG (Science Archive Server) for FITS file retrieval.
4 SELECT TOP 10
5 p.objid, p.ra, p.dec, p.p, p.g, p.r, p.i, p.z,
6 p.run, p.runRef, p.conc, p.field,
7 s.specobjid, s.class, s.z, s.redshift,
8 s.plate, s.hjd, s.fiberid
9 FROM PhotoObj AS p
10 JOIN SpecObj AS s ON s.bestobjid = p.objid
11 WHERE
12 p.ra BETWEEN 0 AND 15.6
13 AND p.dec BETWEEN 0 AND 20
14
```

Output Format: HTML CSV XML JSON OVD Table FITS

Sample Queries

- Basic SQL
- SQL Jujitsu
- Black Hole Mapper
- Milky Way Mapper
- Miscellaneous
- Variability Queries
- General Astronomy
- Galaxies
- Stars
- Quasars
- BOSS
- APOGEE

# Surveys as Astronomical Data Sources

## Example: SDSS

`sdss-access` provides Python utilities for downloading files from the Science Archive Server (SAS). Examples and more details can be found on the official documentation.

<https://www.sdss.org/dr18/software/packages/sdss-access/>

install `sdss-access` with:

```
pip install sdss-access
```

Motivation

Surveys

Services

Using  
`astroquery`

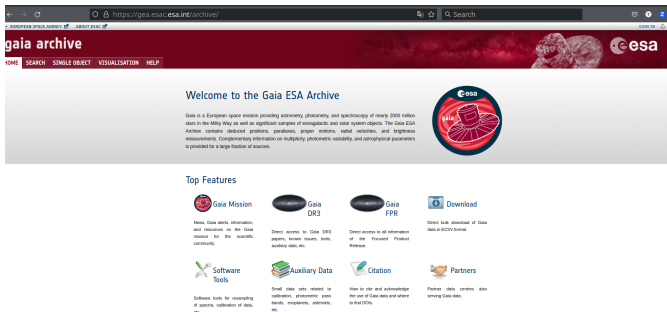
Choosing the  
Right Data  
Source

Outlook

# Surveys as Astronomical Data Sources

Example: Gaia

<https://gea.esac.esa.int/archive/>



The screenshot shows the Gaia ESA Archive website in a web browser. The address bar displays <https://gea.esac.esa.int/archive/>. The website has a red header with the "gaia archive" logo and the ESA logo. Below the header, a navigation bar includes links for HOME, SEARCH, SIMPLE OBJECT, VISUALISATION, and HELP. The main content area features a "Welcome to the Gaia ESA Archive" message, a paragraph describing the mission's goals, and a circular Gaia logo. A "Top Features" section follows, with eight icons and links: Gaia Mission, Gaia DR3, Gaia FPR, Download, Software Tools, Auxiliary Data, Citation, and Partners. Each feature has a brief description of its purpose.

gaia archive

WELCOME SEARCH SIMPLE OBJECT VISUALISATION HELP

Welcome to the Gaia ESA Archive

Gaia is a European space mission providing astrometry, photometry, and spectroscopy of nearly 2000 million stars in the Milky Way as well as significant samples of extragalactic and solar system objects. The Gaia ESA Archive contains dedicated positions, parallaxes, proper motions, radial velocities, and brightness measurements. Complementary information on multiplicity, photometric variability, and astrophysical parameters is provided for a large fraction of sources.

Top Features

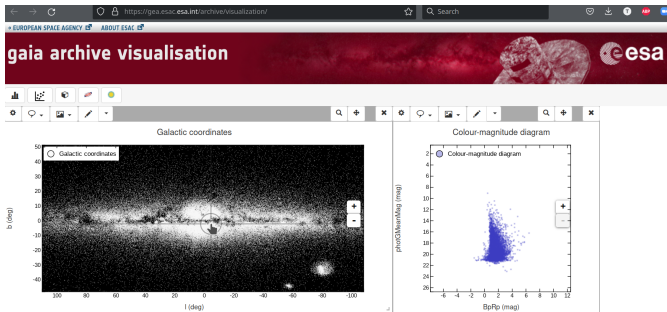
- Gaia Mission**  
News, data alerts, information, and resources on the Gaia mission for the scientific community.
- Gaia DR3**  
Direct access to Gaia DR3 papers, known reports, books, auxiliary data, etc.
- Gaia FPR**  
Direct access to all information of the Forecast Product Reference.
- Download**  
Direct bulk download of Gaia data in ECDF format.
- Software Tools**  
Software tools for resampling of spectra, calibration of data, etc.
- Auxiliary Data**  
Small data sets related to calibration, photometric pass bands, measurements, astrometry, etc.
- Citation**  
How to cite and acknowledge the use of Gaia data and where to find DOIs.
- Partners**  
Partner data centres also serving Gaia data.



# Surveys as Astronomical Data Sources

Example: Gaia

<https://gea.esac.esa.int/archive/visualization>



# Surveys as Astronomical Data Sources

## Example: Gaia

In addition to the bulk download of data, one can write Astronomical Data Query Language (ADQL) queries which are similar to SQL queries.

### **example use case: Cone search sorted by angular separation**

retrieve a sample of filtered sources (brighter than  $G=20.5$  mag and with a parallax measurement) in a circular region centred on the LMC [(R.A., Dec) = (81.28, -69.78) deg] with a search radius of 5 arcmin. The output should be ordered by the angular separation, from small to large.

Target table: `gaiadr3.gaia_source`

Query:

```
SELECT *, DISTANCE(81.28, -69.78, ra, dec) AS ang_sep
FROM gaiadr3.gaia_source
WHERE DISTANCE(81.28, -69.78, ra, dec) < 5./60.
AND phot_g_mean_mag < 20.5
AND parallax IS NOT NULL
ORDER BY ang_sep ASC
```

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# Surveys as Astronomical Data Sources

## Example: Gaia

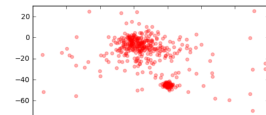
ESA Gaia Archive also provides a Python Astroquery package for easy access to the ESA Gaia Archive: `astroquery.gaia`.

In addition, a generic package is provided for accessing any TAP-compliant service: `astroquery.utils.tap`.

An example Jupyter notebook is provided:

<https://www.cosmos.esa.int/web/gaia-users/archive/use-cases#ClusterAnalysisPythonTutorial>

```
068277898636826938
06747549510136469
06649959804011275
0666478511176576
06518858202869649
041237885940444952
06517885719546219
0489598041317844
0488788131473844
0643621373893792
0643175262098812
0413559425942569
0413559425942569
Length = 30538 rows
```



# Services

With a lot of different surveys existing, a **common scripted interface** to tie all these services together is a good way to make the data more accessible.

Doing so, also provides **reproducibility** of analysis done for publications. A centrally maintained library also safeguards against lost links to data archives, moving some of the responsibility for maintaining long-term reproducibility from each individual researcher to the broader community.

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# CDS Services

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

The Strasbourg astronomical Data Center (CDS) is a data center dedicated to the collection and worldwide distribution of astronomical data and related information. It is located at the Strasbourg Astronomical Observatory, France.

The CDS gives access to many astronomical data such as objects, catalogues, images and spectra.

Several services are available to search for and to get these data. We will look here at two of their main services.

# CDS Services

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

The SIMBAD astronomical database provides basic data, cross-identifications, bibliography and measurements for astronomical objects outside the solar system. SIMBAD can be queried by object name, coordinates and various criteria. Lists of objects and scripts can be submitted. Links to some other on-line services are also provided.

The **SIMBAD** database currently contains information for about 13,000,000 astronomical objects (stars, galaxies, planetary nebulae, clusters, novae and supernovae, etc.).

The only astronomical objects specifically excluded from SIMBAD are Solar System bodies (planets, satellites, asteroids, comets)

# CDS Services

`https://simbad.unistra.fr/simbad/`

The screenshot displays the SIMBAD web interface for script execution. At the top, the URL `https://simbad.unistra.fr/simbad/sim-fscript` is shown in the browser's address bar. Below the navigation bar, the page title is "SIMBAD: Script execution". A row of buttons includes "other query", "Identifier query", "Coordinate query", "Criteria query", "Reference query", "Basic query", "Script submission", "TAP", "Output options", and "Help". The "Script submission" button is highlighted.

Below the buttons, a section titled "Enter the name of an ASCII file containing a script:" contains a "Browse..." button (with "no file selected" text), a "submit file" button, a "clear" button, and checkboxes for "file output" and "compressed".

The main area is titled "Type your script in:" and features a large text input field for the script, a "submit script" button, and "clear", "file output", and "compressed" checkboxes. To the right of the input field is a "Short reminder of the script commands. A more complete description can be found behind the help anchor".

The reminder lists the following commands:

- control:**
- # comment**: a comment line must have a '#' as its first char
- output outfile=action...**: output definition. 'outfile' can be 'console', 'error' or 'script'
- 'action' can be 'off' or 'merge(line header)'**: (merge into the data output)
- result type**: define the kind of result (full|oid|ident|count)
- votable [name] {fieldlist}**: define a votable output with a fieldlist
- votable open [name]**: use the defined votable output for the next queries
- votable close**: close the current votable
- format {obj|ref} [name] "..."**: define the output formats for objects or references
- format name**: set the named output format as current
- format {obj|ref} default**: reset the output formats for obj or ref to the default
- format display [name]**: display all format or the one defined by its name
- set limit #rows**: limits the number of returned rows (0 removes any limitation)
- set radius val[d|m|s]**: radius for coordinates and around queries (Ex: 1.5d, 10m, 3s)
- set frame name**: input frame for coord. queries (ICRS, FK5, FK4, GAL, SGAL, ECL)
- set equinox val**: Equinox value for coord. queries (Ex: 2006.5)
- set epoch val**: Epoch value for coord. queries (Ex: B1950, J2000)
- default for coord. queries: radius=10m, frame=ICRS, equi=2000.0, epc**
- without a value, a set xxx command resets the default**
- echo [-n] text**: simple text to display on the console. option -n: no newline at the end
- echodata [-n] text**: 'echodata' displays on the data output. option -n: no newline at the end
- quit**: Allows to quit before the end of a script

# CDS Services

The **VizieR** service provides mostly catalogs associated with papers or authors.

<https://vizier.cds.unistra.fr/viz-bin/VizieR>

The screenshot shows the VizieR web interface. At the top, there's a navigation bar with links to Portal, Simbad, VizieR, Aladin, X-Match, and other services. The main heading is "VizieR". Below it, a message states: "The VizieR service is now hosted by CDS domain (cds.unistra.fr). Please, modify your configuration for the new domain." The interface is divided into several sections: "Search Criteria" on the left with preferences like "max: 50" and "HTML Table"; "Find catalogs among 24534 available" with a search bar and "Expand search" button; "Catalog, author's name, word(s) from title, description, etc." with examples; "Search for catalogs by column descriptions (UCD)" and "Search for catalogs containing additional data"; "Search by Position across 27563 tables" with fields for "Target Name (resolved by Seagame) or Position" and "Target dimension" (set to 2 arcmin); and a "Wavelength Mission Astronomy" table. At the bottom, there's a "Tools related to VizieR" section with links to various services.

Wavelength	Mission	Astronomy
Radio	AKARI	Abundances
Millimeter	ANS	Ages
IR	ASCA	AGN
optical	BeppoSAX	Associations
UV	Cassini-Huygens	Asteroseismology
EUV	CGRO	Atomic_Data
X-ray	Chandra	Binaries:catclysmic

## Tools related to VizieR

- [Catalogue collection](#) : Search VizieR catalogues available via various services (FTP, VizieR, TAP, ...)
- [CDS Portal](#) : Access CDS data including VizieR, Simbad and Aladin using the CDS portal
- [Spectra, images in VizieR](#) : Search Spectra, images in VizieR
- [Photometry viewer](#) : Plot photometry (sed) including all VizieR
- [TAP VizieR](#) : query VizieR using ADQL (a SQL extension dedicated for astronomy)
- [CDS cross-match service](#) : fast cross-identification between any 2 tables, including VizieR catalogues, SIMBAD



# CDS Services

An extensive list of Python tutorials on how to interface with the CDS services is available:

<https://cds-astro.github.io/tutorials/intro.html>

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# Community brokers

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

The newest generation of large survey telescopes such as the Vera C. Rubin Observatory (which will conduct the 10-year Legacy Survey of Space and Time (LSST)) are designed to produce large streams of transient alerts. They will be processed, annotated and classified by a new generation of astronomical software for time-domain surveys, called **alert brokers**.

Brokers are currently developed tested with the Zwicky Transient Facility (ZTF). Usually, brokers provide a website as well as a Python interface.

# Community brokers

Example: The ANTARES Broker

<https://antares.noirlab.edu/>

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

The screenshot displays the ANTARES web interface. The top navigation bar includes 'Explore', 'Favorites', 'Filters', 'Tags', 'Watch Lists', 'Catalogs', 'Pipeline', and 'Properties'. A search bar is located on the right. The main content area shows a table of astronomical objects with the following columns: ID, RA, Dec, Latest Mag, Brightest Mag, # Alerts, Latest Alert, First Alert, and Actions. The table lists several objects, including ANT2020g8y6, ANT2020g8y6, ANT2020g8y6, ANT2020g8y6, ANT2020g8y6, ANT2020g8y6, ANT2020g8y6, ANT2020g8y6, ANT2020g8y6, and ANT2020g8y6. The left sidebar contains filters for 'Latest Alert Within', 'First Alert Within', 'Number of Measurements', 'Cone Search', and 'Catalogs'. The 'Catalogs' section lists various astronomical catalogs such as 'galaxy', 'zmass', 'allstars', 'bright', 'sdss', 'gaia', 'psa', 'var', and 'wga'.

ID	RA	Dec	Latest Mag	Brightest Mag	# Alerts	Latest Alert	First Alert	Actions
ANT2020g8y6	113.49	3.88	17.47	15.49	122	2023-12-15 13:44:48	2018-12-28 10:20:18	...
ANT2020g8y6	113.59	3.68	16.40	15.98	436	2023-12-15 13:44:48	2018-11-28 09:05:51	...
ANT2020g8y6	112.94	5.08	17.13	16.34	303	2023-12-15 13:44:48	2018-11-28 10:59:51	...
ANT2020g8y6	112.90	4.74	17.96	16.06	5	2023-12-15 13:44:48	2021-02-10 05:04:10	...
ANT2020g8y6	112.65	7.86	18.29	17.65	15	2023-12-15 13:44:48	2019-09-30 11:29:21	...
ANT2020g8y6	112.21	7.45	17.70	17.27	262	2023-12-15 13:44:48	2018-11-28 09:05:51	...
ANT2020g8y6	110.28	4.85	16.12	14.82	116	2023-12-15 13:44:48	2019-10-07 11:28:01	...
ANT2020g8y6	110.32	4.79	17.92	16.86	46	2023-12-15 13:44:48	2018-12-28 10:28:39	...
ANT2020g8y6	110.68	6.98	16.68	14.74	203	2023-12-15 13:44:48	2018-11-28 10:59:51	...
ANT2020g8y6	110.16	3.20	17.66	16.11	28	2023-12-15 13:44:48	2018-12-17 11:07:48	...

# Community brokers

## Example: The ALeRCE Broker

<http://alerce.science/>

<https://github.com/alercebrokebroker/usecases>

Motivation

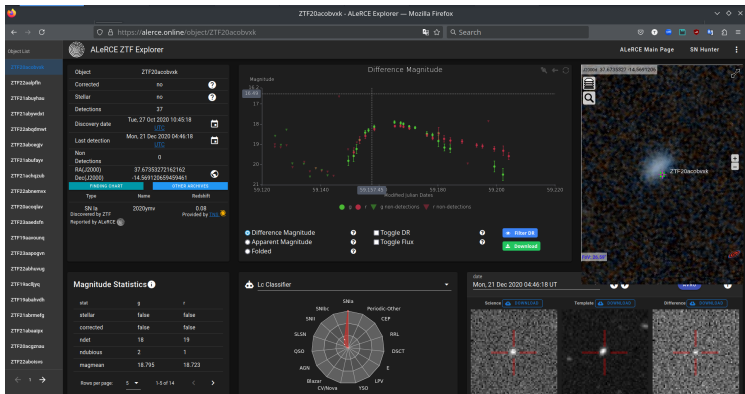
Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook



# Community brokers

Brokers will show data from LSST as soon as LSST will deliver the alert stream.

This will be from approx. mid 2025 on.

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# What is astroquery?

**Astroquery** as part of **astropy** is a set of tools for querying astronomical web forms and databases.

In addition to providing access to astronomical databases, **astroquery** provides data such as molecular and atomic properties, e.g. by the NIST Atomic Spectra Database, bibliographic databases such as the NASA Astrophysics Data System (ADS), or services that are computationally intensive or require regular updates, like solar system ephemerides provided by services like JPL HORIZONS, or the Minor Planet Center.

<https://github.com/astropy/astroquery>  
<https://astroquery.readthedocs.io/>



Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# What is astroquery?

These tools are built on the Python `requests` package, which is used to make HTTP requests, and `astropy`, which provides most of the data parsing functionality. `astroquery` modules generally attempt to replicate the web page interface provided by a given service as closely as possible, making the transition from browser-based to command-line interaction easy.

The complete documentation for `astroquery` can be found at <http://astroquery.readthedocs.io/>.

Motivation

Surveys

Services

Using  
`astroquery`

Choosing the  
Right Data  
Source

Outlook

# Using astroquery

All astroquery modules are supposed to follow the same API. In its simplest form, the API involves queries based on coordinates or object names. Some simple examples, using SIMBAD:

```
from astroquery.simbad import Simbad

result_table = Simbad.query_object("m1")

print(result_table)
```

MAIN_ID	RA	DEC	...	COO_BIBCODE	SCRIPT_NUMBER_ID
	"h:m:s"	"d:m:s"	...		
-----	-----	-----	...	-----	-----
M 1	05 34 30.9	+22 00 53	...	1995AuJPh..48..143S	1

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook



# Using astroquery

All query tools allow coordinate-based queries:

```
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)
print(result_table)
```

MAIN_ID	RA	...	COO_BIBCODE	SCRIPT_NUM
	"h:m:s"	...		
NAME Ori Region	05 35 17.30	...		
...	...	...	...	
2MASS J05353573-0525256	05 35 35.7755	...	2020yCat.1350....OG	
V* V2114 Ori	05 35 01.6720	...	2020yCat.1350....OG	

Length = 3272 rows

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# Using astroquery

The following is an alphabetic **selection of available services** through astroquery; the full list can be found in the documentation:

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

ALMA Queries (`astroquery.alma`)

Atomic Line List (`astroquery.atomic`)

Besancon Queries (`astroquery.besancon`)

CADC (`astroquery.cadc`)

CASDA Queries (`astroquery.casda`)

Cologne Database for Molecular Spectroscopy (CDMS) Queries  
(`astroquery.linelists.cdms`)

DACE (`astroquery.dace`)

ESA Herschel Science Archive (`astroquery.esa.hsa`)

ESA HST Archive (`astroquery.esa.hubble`)

ESA ISO Archive (`astroquery.esa.iso`)

ESA JWST Archive (`astroquery.esa.jwst`)

ESA XMM-Newton Archive (`astroquery.esa.xmm_newton`)

# Choosing the Right Data Source

Which **survey** you use is mostly determined by the science you are doing.

How you **access** the data is mostly determined by the amount of data.

Often, a survey provides multiple ways for accessing the same data.

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

# Chosing the Right Data Source

## **Small data sets, exploratory:**

Explore them directly on the website, e.g. clicking through images, plots.

### **example:**

- Using the broker website.
- Using small, older surveys with digitized data.

Motivation

Surveys

Services

Using  
astroquery

Chosing the  
Right Data  
Source

Outlook

# Choosing the Right Data Source

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

## **Small data sets, exploratory:**

Explore them directly on the website, e.g. clicking through images, plots.

### **example:**

- Using the broker website.
- Using small, older surveys with digitized data.

## **Medium-sized data sets:**

Download survey data ("bulk download") and process locally.

### **example:**

- Downloading a catalog of variable stars in Pan-STARRS.
- Downloading a sector of TESS data to search for specific objects.

# Choosing the Right Data Source

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

## Small data sets, exploratory:

Explore them directly on the website, e.g. clicking through images, plots.

### example:

- Using the broker website.
- Using small, older surveys with digitized data.

## Medium-sized data sets:

Download survey data ("bulk download") and process locally.

### example:

- Downloading a catalog of variable stars in Pan-STARRS.
- Downloading a sector of TESS data to search for specific objects.

## Very large data sets:

Access data using Python packages. Select data using e.g. SQL/ADQL, download only selected data. Process locally or in the cloud.

### example:

- Accessing SDSS, Gaia.
- Python broker interfaces, downloading only specific alerts.

# Summary & Outlook

Motivation

Surveys

Services

Using  
astroquery

Choosing the  
Right Data  
Source

Outlook

We have seen ways on how to access astronomical databases from within Python.

In the next lecture we will continue with **Astronomical Packages** which provide us with functionality such as coordinate conversion, data reduction (photometry, spectroscopy) and other more specialized calculations in astronomy.