New Python Developments to Access CDS Services

astroquery.cds package & MOCPy improvements

Matthieu Baumann & Thomas Boch
Centre de Données Astronomiques de Strasbourg
matthieu.baumann@astro.unistra.fr

Abstract

We will present recent developments made in the frame of the ASTERICS project and aimed at providing Python interface to CDS services and Virtual Observatory standards. Special care has been taken to integrate these developments into the existing astropy/astroquery environment.
A new astroquery.cds module allows one to retrieve image or catalogue datasets available in a given region of the sky described by a MOC (Multi Order Coverage map) object. Datasets can also be filtered through additional constraints on their metadata.
The MOCPy library has been upgraded: performance has been greatly improved, unit tests and continuous integration have been added, and the integration of the core code into the astropy.regions module is under way. We have also added an experimental support for creation and manipulation of T-MOCs which describe the temporal coverage of a data collection.

I MOCPy [4]: a Library Handling the Creation and Manipulation of MOCs

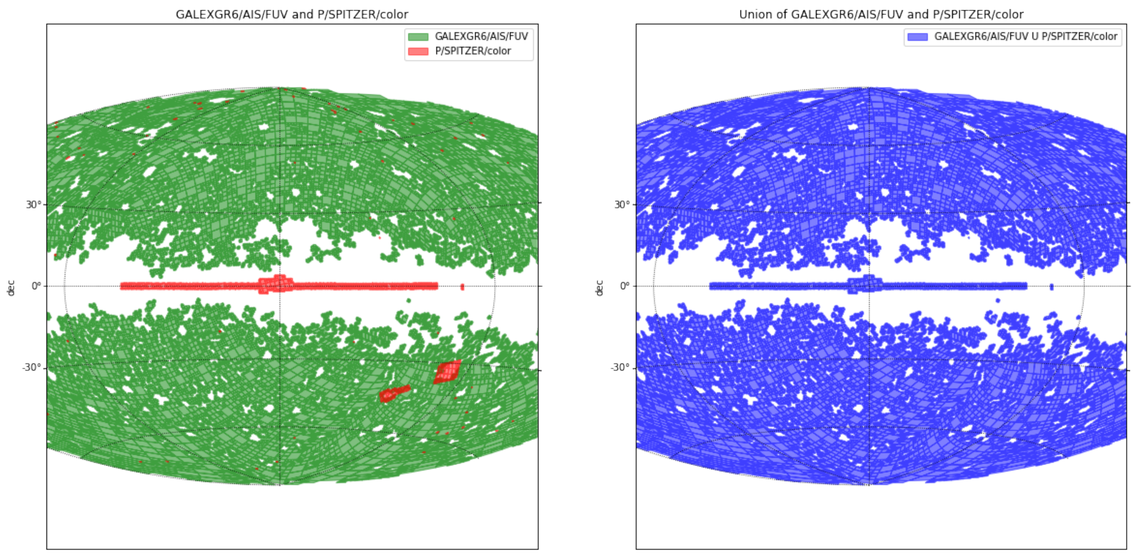
New features and improvements have been added to the library:

- MOCPy [4] has been optimized and tends to use numpy's broadcasting feature as much as possible. Creating a MOC from a list of `astropy.coordinates.SkyCoord` is a lot faster thanks to the vectorization involved when operations are directly done on numpy arrays.
  The following code shows the implementation of `from_lonlat` responsible for creating a MOC from lon and lat astropy quantities at a given `order`.

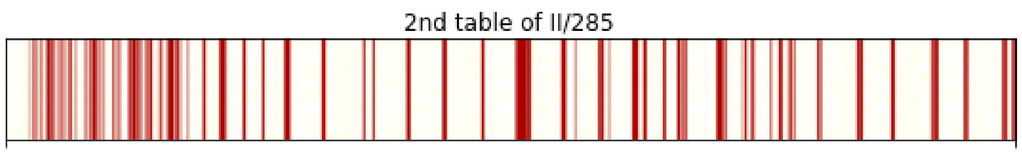
```
In [1]: from astropy_healpix import HEALPix
        # lon and lat are astropy quantities
        hp = HEALPix(nside=(1 << order), order='nested')
        ipix = hp.lonlat_to_healpix(lon, lat)

        shift = 2 * (29 - order)
        intervals = np.vstack((ipix << shift, (ipix + 1) << shift)).T
```

  This code:
  - Uses `astropy-healpix` to get the HEALPix cells where the (lon, lat) coordinates are located.
  - Build a $N \times 2$ numpy array storing the intervals of the HEALPix cells at a given order.

  No Python loops over the quantities are involved here as it is encouraged to perform operations directly on numpy arrays.
- Dependencies to healpy have been removed. We now use astropy-healpix and therefore have changed the licence of MOCPy [4] from GPL to BSD-3.
- A new `serialize` method has been added, taking an optional format argument that can be set to fits or json.
- New methods `fill` and `perimeter` have been implemented. These methods are responsible for plotting the MOC (resp. its perimeter) on a `matplotlib.axe.Axes` using a projection defined by an `astropy.wcs.WCS` object.



- A new `TMOC` class handles the creation and manipulation of temporal MOCs. A `from_times` method creates a T-MOC object from an `astropy.time.Time` object. As for the spatial MOCs, it is possible to `serialize` a T-MOC, compute the intersection, union, difference between several T-MOCs as well as use them to filter an `astropy.time.Time` object.




```
First observation: 1978-05-10 20:09:28.672
Last observation: 2004-04-22 16:56:36.350
Total duration: 227.424 jd
Max order: 14
```

II astroquery.cds [1]: a New Module for Retrieving Data Collections Based on Region and/or Meta-data Queries

astroquery.cds [1] has been merged into the master branch of astroquery in July the 23th and will be available for its next release (v0.3.9). This module requests the CDS MOCServer, a server storing MOCs and meta-data of $\simeq$ 20000 data collections. This package offers two methods (see the module's documentation [1] for more details):

- `query_region` retrieves the collections having their observations in a specific region. Regions can be expressed as `regions.CircleSkyRegion/PolygonSkyRegion` or `mocpy.MOC` objects.
- `find_datasets` retrieves the collections based on a constraint on their meta-data.

These two methods return by default an `astropy.table.Table` containing the meta-data of one collection per row. An optional argument `return_moc=True` can be used to directly retrieve the MOC (a `mocpy.MOC` object) of the matching collections.

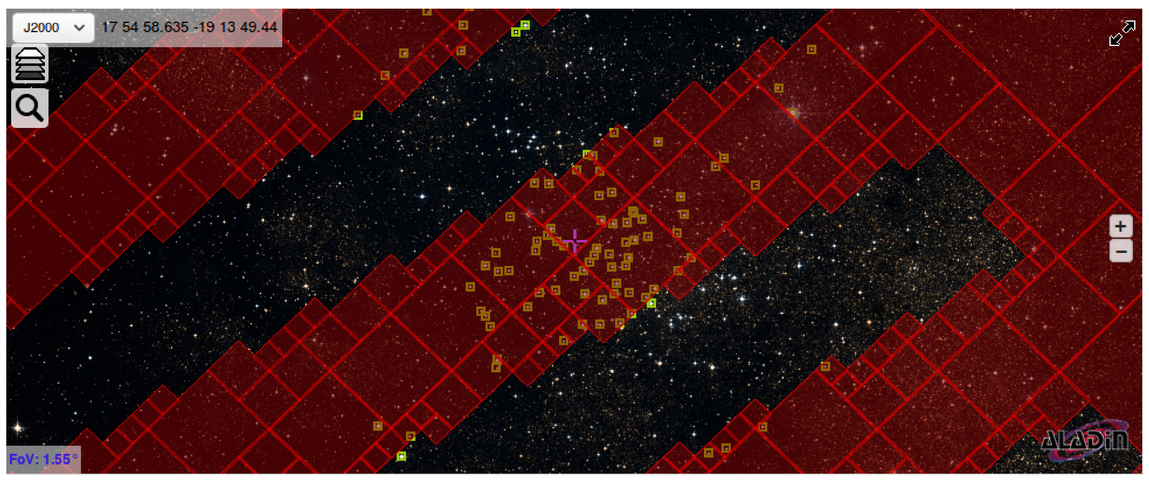
Below is an example of an `astropy.table.Table` returned by `query_region` and filtered to select only the vizier tables having between 75000 and 100000 sources. The meta-data shown here are `obs_id`, `obs_title`, `dataproduct_type` and `cs_service_url`. For a list of all the possible meta-data returned by the `cds` module, please refer to the page 18 of the HiPS IVOA paper [3].

obs_id	obs_title	dataproduct_type	cs_service_url
str28	str91	object	str91
I/208/ppm3	The 90000 stars Supplement to the PPM Catalogue (Roeser+, 1994) (ppm3)	catalog	http://vizier.u-strasbg.fr/viz-bin/votable/-A?-source=I%2F208%2Fppm3&
I/237/catalog	The Washington Visual Double Star Catalog, 1996.0 (Worley+, 1996) (catalog)	catalog	http://vizier.u-strasbg.fr/viz-bin/votable/-A?-source=I%2F237%2Fcatalog&
I/276/catalog	Tycho Double Star Catalogue (TDSC) (Fabricius+ 2002) (catalog)	catalog	http://vizier.u-strasbg.fr/viz-bin/votable/-A?-source=I%2F276%2Fcatalog&
II/168/ubvmeans	Homogeneous Means in the UBV System (Mermilliod 1991) (ubvmeans)	catalog	http://vizier.u-strasbg.fr/viz-bin/votable/-A?-source=II%2F168%2Fubvmeans&

III State of the Art of the CDS Python Tools

The following image results from a notebook script combining different Python packages, most of them being developed by the CDS team through the past years. It is available on the cds-astro [2] github repository as an example for astronomers. This script:

1. Retrieves two MOCs from the MOCServer (astroquery.cds [1]).
2. Computes their intersection (MOCPy [4]) and shows the resulting MOC on an aladin-lite view (ipyaladin).
3. Searches for a vizier table in optical regime having some observations in this region (astroquery.cds [1]).
4. Retrieves the table using astroquery.vizier.
5. Filters the table to only keep the observations lying in the MOC (MOCPy [4]) and adds the filtered table to the aladin view (ipyaladin).



IV Future Improvements

- MOCPy [4] is currently being integrated into astropy/regions. New classes, `MOCSkyRegion` and `MOCPixelRegion` will be implemented. `MOCSkyRegion` is the equivalent of the `mocpy.MOC` class, therefore it will contain all its features (serialization, intersection, ...). A `MOCPixelRegion` is a MOC sky region projeted on an astropy WCS object.
- `astroquery.cds.query_region` will be upgraded to accept `MOCSkyRegion` objects.
- The `query_region` methods of both `astroquery.simbad` and `astroquery.vizier` should accept `MOCSkyRegion` too so that Simbad and Vizier tables can be filtered by MOCs.

References

[1] Matthieu Baumann. astroquery.cds documentation page. https://astroquery.readthedocs.io/en/latest/cds/cds.html, 2018.

[2] Matthieu Baumann. Notebook example illustrating the state of the art of the CDS Python tools. https://github.com/cds-astro/ADASS-IVOA18, 2018.

[3] Pierre Fernique, Mark Allen, Thomas Boch, Tom Donaldson, Daniel Durand, Ken Ebisawa, Laurent Michel, Jesus Salgado, and Felix Stoehr. Hips–hierarchical progressive survey version 1.0. 2017.

[4] Thomas Boch Matthieu Baumann. Python library to easily create and manipulate MOCs (Multi-Order Coverage maps) . https://github.com/cds-astro/mocpy, 2015-.

ADASS XXVIII - November 11-15, 2018 in College Park, Maryland, USA – CDS – Observatoire astronomique de Strasbourg