ESA Datalabs with Pandas: Creating 126 Million Source Cutouts

The upcoming release of ESA Datalabs will allow a direct interface between the user and the large data repositories of space telescopes such as the Hubble and James Webb Space Telescopes. We will demonstrate how to leverage this interface to efficiently create cutouts of sources from the Hubble Source Catalogue using the Python package Pandas. To do this before ESA Datalabs, the user would have to use ASQL commands in TAP services, AstroQuery or other network dependent methods to get coordinates of sources and then download such cutouts using online cutout services or even download the FITS files containing them. While quick in theory, these methods are often plagued by network throttling and/or long download times. Downloading a single FITS file can be a lengthy procedure, taking minutes in some cases. With ESA Datalabs providing direct access to the archives and the source files the user has unparalleled control and efficiency in creating source cutouts. We will demonstrate how we used ESA Datalabs to create 126 million source cutouts - all the extended sources in the Hubble Source Catalogue - to use with machine learning to find galaxies undergoing interaction. The result of this project was a catalogue of 21,765 interacting galaxies, a factor of ten larger than any previously existing interacting galaxy catalogue. 14,404 of these - approximately 65% - had no prior reference in SIMBAD or the NASA Extragalactic Database. The process of creating the cutouts alone could easily have taken far longer than the end-to-end project including creating the training set to making predictions on all sources. Thanks to ESA Datalabs, this project was completed in a total of three months.