

Exploring big data efficiently with the Astro Data Lab science platform

ADASS XXXIII

Tucson, AZ, November 5-9, 2023

Presenters:

Alice Jacques, Stéphanie Juneau, Robert Nikutta, David Herrera, Yumi Choi

Description of main topic

In the era of data-intensive astronomy the community needs to acquire skills to handle increasingly larger and more complex datasets, and to access high-performance computing and analysis tools. In this tutorial we will teach participants how to use data-proximate science platforms to conduct astronomy research. Using the Astro Data Lab science platform and the new SPARCL (SPectra Analysis and Retrievable Catalog Lab) service for spectroscopy, participants will first learn how to find documentation, information about all of Astro Data Lab's data holdings of over 100 TB of wide-field survey catalogs, 2.5 PB of imagery, and millions of spectra from DESI and SDSS, and how to access help from the Astro Data Lab team. We will then teach the group in an interactive mode how to use various data services and analysis tools at Data Lab, including how to crossmatch tables, build and submit catalog queries, search for images and create cutouts, search for and download spectra, and how to use the Astro Data Lab Jupyter notebook server. The participants will execute and modify a number of science-case example notebooks from various domains of astronomy focusing on data analysis. The tutorial will also make use of some amenities on science platforms, including remote file storage and remote user-owned database tables.

Primary learning objectives

- Construct SQL queries to query large datasets through a dedicated Jupyter Notebook server, web-interface, and command-line interface
- Use Astro Data Lab's X-match web service to crossmatch datasets
- Discover and query for Sloan Digital Sky Survey (SDSS) and Dark Energy Spectroscopic Instrument (DESI) spectra with the new SPARCL tool
- Create plots with the data obtained to realize the graphical and visualization capabilities within our notebooks
- Get image cutouts for a set of objects using an image discovery service and a cutout tool

Detailed tutorial structure

- (10 min) Brief introductory presentation on what the Astro Data Lab science platform is and how it can be utilized to explore, discover, and analyze data easily and efficiently. Slide deck (PDF) is available for preview here:
 - github.com/astro-datalab/Tutorial-ADASS-2023/blob/main/ADASS-2023-DataLab-tutorial.pdf

- (45 min) Hands-on tutorial of using Astro Data Lab
 - Registering for an account: <https://datalab.noirlab.edu/account/register.html>
 - *Ideally done in advance*
 - Documentation / user manual: <https://datalab.noirlab.edu/docs/manual/>
 - Show TOC and search function
 - **Exercise:** find the answer to “How to use TOPCAT with the Astro Data Lab?”
 - Helpdesk: <https://datalab.noirlab.edu/help>
 - Show how to search for keywords in previous answers
 - Show how to ask a new question
 - Database table browser: <https://datalab.noirlab.edu/query.php>
 - Show schemas, tables, columns (names, dtypes, descriptions)
 - Query tool
 - **Exercise:** execute a SQL query and save the table to MyDB
 - X-match web interface service: <https://datalab.noirlab.edu/xmatch.php>
 - Crossmatch two datasets (based on new user table in MyDB, or optionally users can bring their own data table)
 - **Exercise:** use the X-match tool to crossmatch previous MyDB table and another Data Lab dataset
 - All-sky viewer: <https://datalab.noirlab.edu/discovery.php>
 - Jupyter notebook server: <https://cloud.datalab.noirlab.edu>
- (3 min) Break
- (7 min) Intro to SPARCL slides and website tour: <https://astrosparcl.datalab.noirlab.edu/>
- (45 min) Hands-on exercises using available Jupyter notebooks as a starting point (available in this repository: <https://github.com/astro-datalab/Tutorial-ADASS-2023>):
 - Getting started with Astro Data Lab
 - Comparison of DESI and SDSS spectra using SPARCL
 - Detecting the Hydra II dwarf galaxy in SMASH DR1 (plus exercise)
- (10 min) Q&A

Participants will need

- A laptop
- Connection to the internet
- A Data Lab account (sign up here: <https://datalab.noirlab.edu/account/register.html>)
- Optionally: participants can work with their own data by preparing a CSV data table with at least RA, Dec coordinates for crossmatching (the tutorial will show how to generate such a table)

Infrastructure requirements

- Wifi
- Projector/screen
- Cable for showing laptop screen on projector screen
- Slack channel for this tutorial

For pre-tutorial questions/support, please email datalab@noirlab.edu