# Exploring big data efficiently with the Astro Data Lab science platform

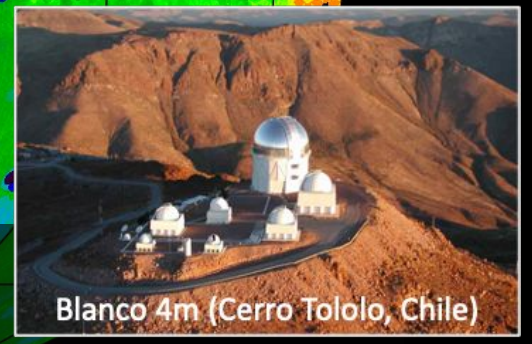
*Astro Data Lab Team <datalab@noirlab.edu>*

NSF's NOIRLab

Tutorial at ADASS XXXIII, November 2023, Tucson/AZ

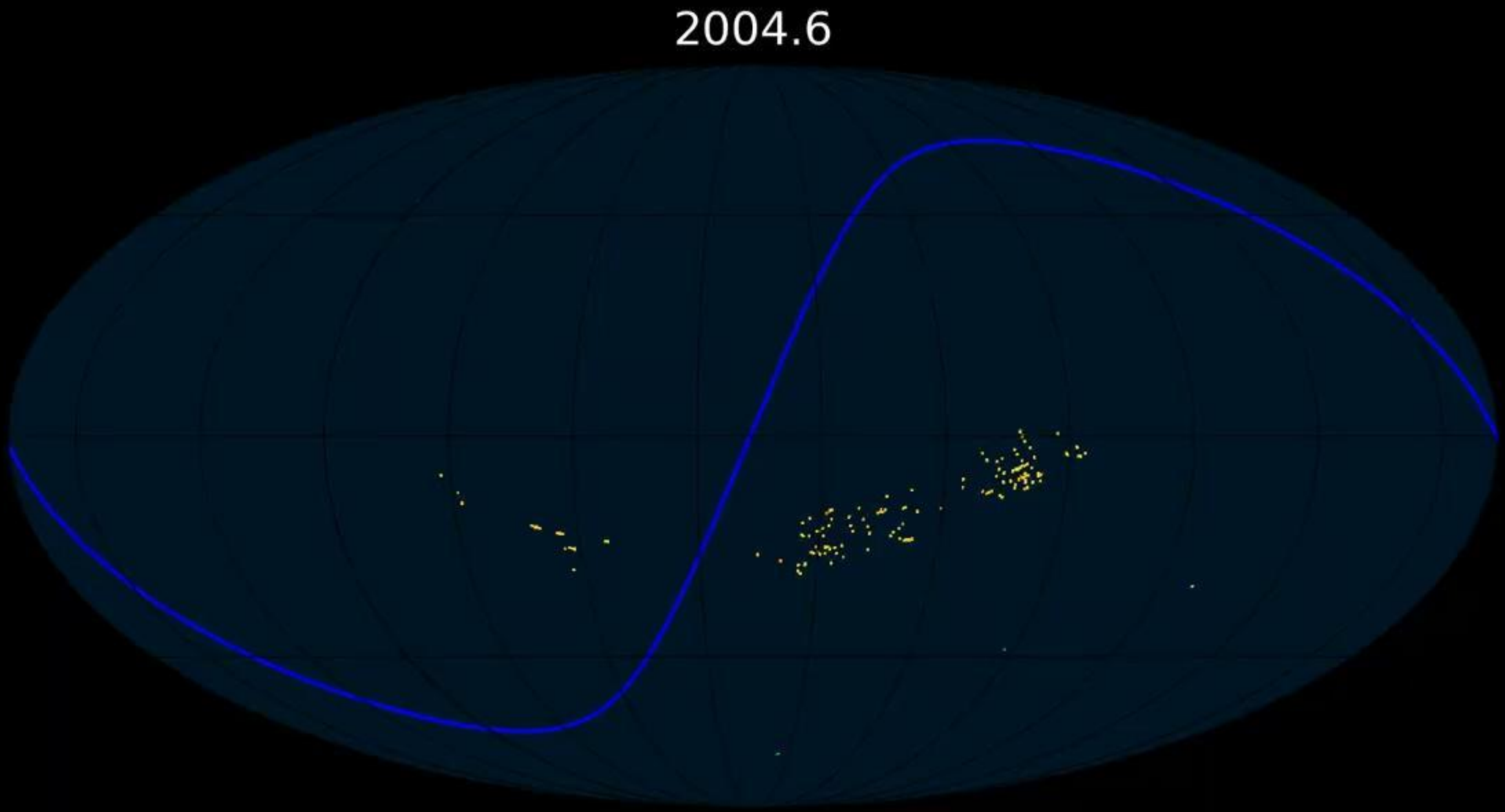
# Introduction to the Astro Data Lab science platform

# The Astro Data Lab


## Mission

Efficient exploration and analysis of large astronomy datasets with an emphasis on NOIRLab wide-field 4-meter telescopes

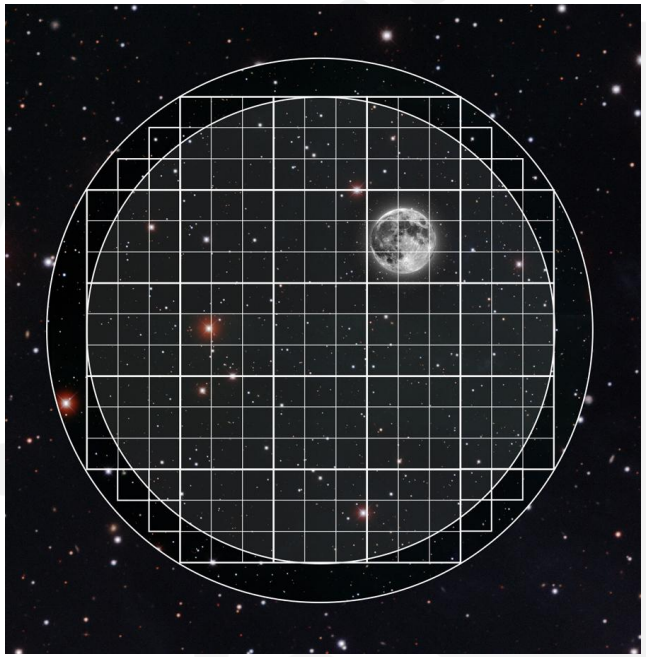



Total exposure time on NOIRLab 4-m class telescopes

# Wide-field cameras & data avalanche

**Dark Energy Camera**
570 Mpix


**Rubin Camera**
3200 Mpix

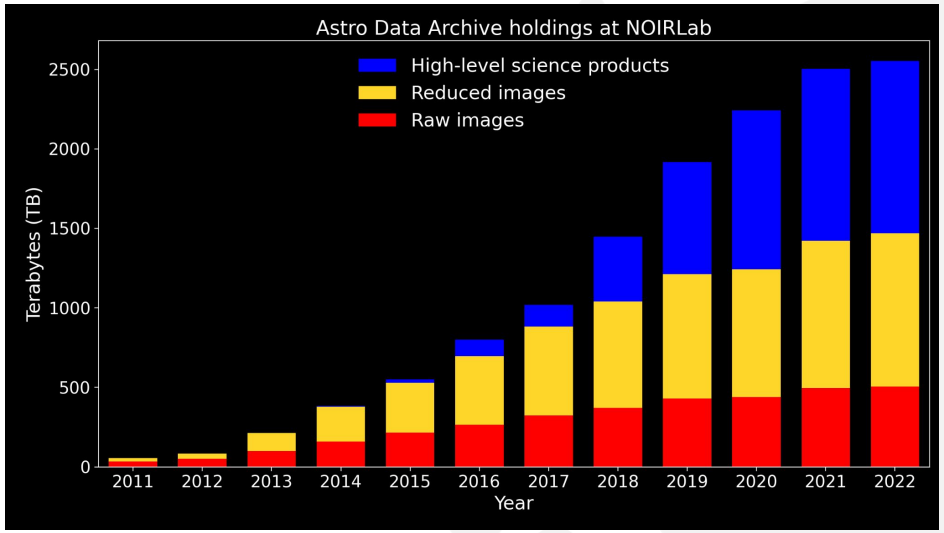


**Rubin Camera**
40x Moon area

# Astro Data Lab Science Platform

## Pre-installed software/tools/tutorials co-located with data

- Web services (website)
- JupyterLab (notebook) server
- Large variety of Python packages




## Rich variety & volume of data

- Images (2.5 PB)
- Catalogs (175 TB) in databases
- Spectra (40+ M) from ground-based & space-based observatories

## Astronomers/students can

- Create a user account, log on
- Use our services for their *entire* analysis (directly from their browser or install a command-line package)

Low barrier of entry to powerful tools

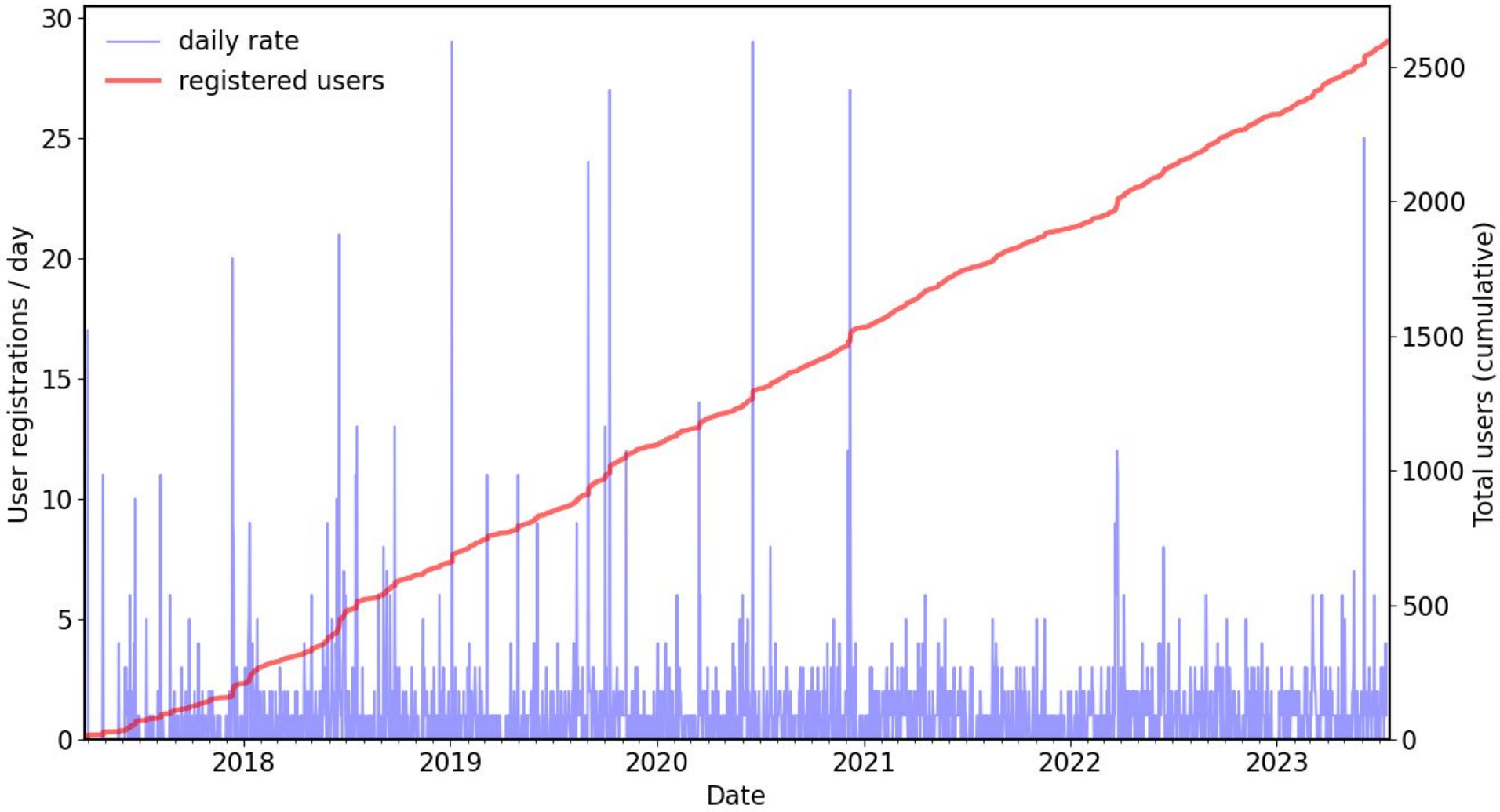
Access to big data

Open & inclusive Community oriented

# Some datasets hosted at DL

| Dataset | Number of objects or measurements | Notes | |
|---|---|---|---|
| DES DR 1 & 2 | 691M objects in DR 2 | photometry | **NOIRLab telescopes** |
| DECam Legacy Surveys DR 3 - 10 | 3.14B objects in DR 10 | photometry, targeting for DESI | |
| NOIRLab Source Catalog DR 1 & 2 | 3.9B objects, 68B measurements in DR 2 | homogeneous reduction across 3 cameras | |
| SMASH DR 1 & 2, DELVE DR 1 & 2 | 2.5B objects in DELVE DR 2 | photometry | |
| PHAT v2 | 117M objects | Andromeda | |
| DESI EDR, SPARCL (metadata) | 2M spectra in DESI EDR | spectroscopy | |
| Gaia DR 1, 2, EDR3, 3 | 1.8B objects in DR 3 | astrometry | **external** |
| AllWISE, unWISE, CatWISE2020 | 2.2B objects in unWISE | IR | |
| VHS DR 5 | 1.4B objects | photometry (S hemisphere) | |
| UKIDSS DR 11+ | 1.2B objects | photometry (N hemisphere) | |
| Buzzard DR 1, LSST SIM DR 2 | 12.6B objects in LSST SIM DR 2 | simulated datasets | **simulated** |

# Growing user community

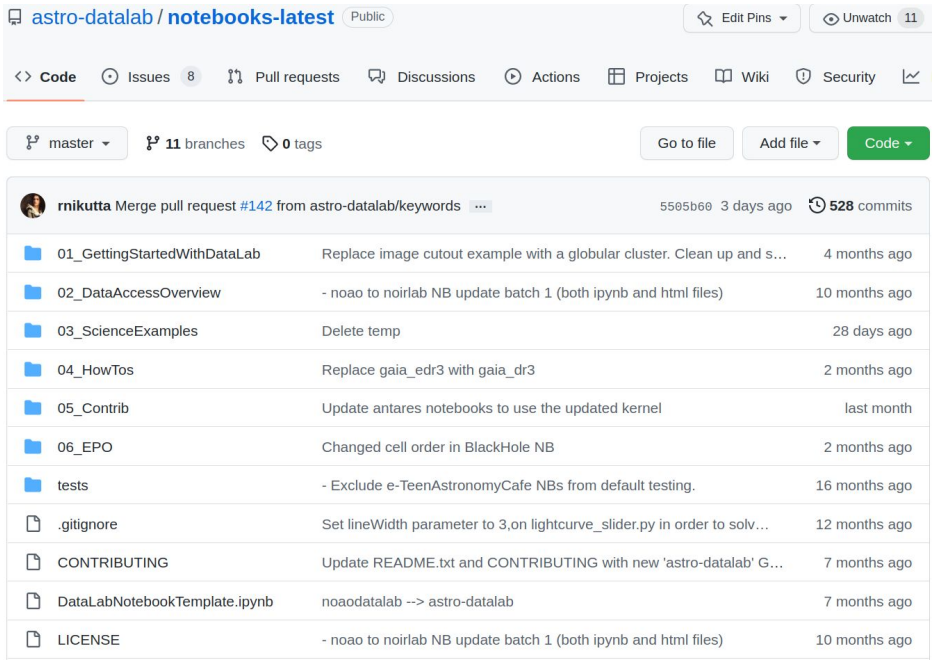



As of July 14, 2023

# Data services & tools

| Action | Modes | Notes |
|---|---|---|
| Authentication | web, client/API, cmdline | log on to access user-specific functions |
| Sky exploration | Aladdin Lite | In-house and external all-sky maps |
| Catalog query | web form, client/API, cmdline | SQL/ADQL, sync/async |
| Cross-matching | web, client/API, cmdline | Also hosting fast pre-crossmatched tables |
| Query results to.. | client, VOSpace, MyDB | Format conversions on the fly |
| Spectro access | client/API | SPARCL |
| Remote user file storage | client/API, cmdline | VOSpace |
| Remote user DB | client/API, cmdline | MyDB |
| Image search & cutout | client/API, web | SIA |
| Analysis (with all of Python) | web, local | Jupyter notebook server |

# Learning & Documentation



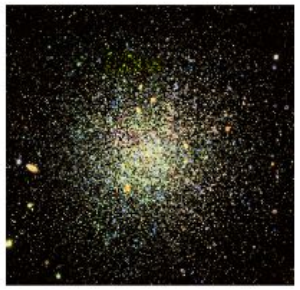

**Curated Jupyter notebook collection**
Intros, How-Tos, Science Cases, EPO, Contributed…


**User Manual & API docs**

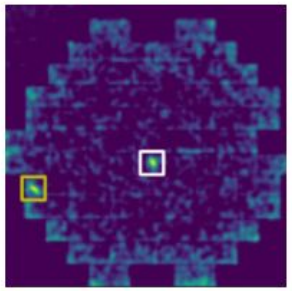

**Helpdesk & FAQ**

# Discovery Ready Science Notebooks

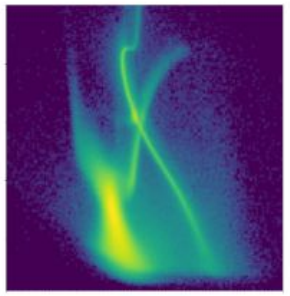
*Astro Data Lab Notebook Gallery*



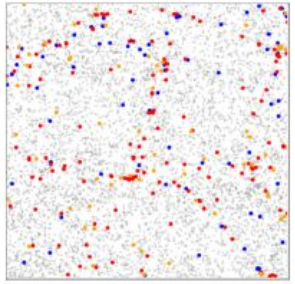
**Getting Started with Data Lab**
Learn the basics such as importing modules, sending a database query, and using the Simple Image Access (SIA) service to create image cutouts.



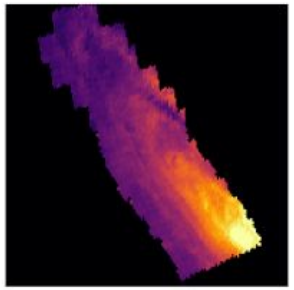
**Dwarf Galaxies in the SMASH survey**
Discover the ultrafaint Hydra II dwarf galaxy in the SMASH DECam survey based on spatial overdensities of blue stars with a detection algorithm.



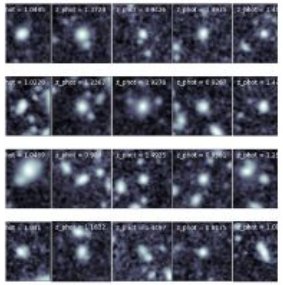
**Exploring SMASH DR2**
The 480 square degree SMASH DECam survey of the Magellanic Clouds and their periphery contains a wealth of objects, including this capture of the SMC with 47 Tuc in the foreground.



**Large-Scale Structure of the Universe**
Investigate cosmic filaments and clusters of galaxies, pan around an interactive sky viewer, combining spectroscopy and DESI pre-imaging.



**Fun with PHAT**
Visualize the 100 million+ stars in the Andromeda Galaxy captured by the Panchromatic Hubble Andromeda Treasury (PHAT).



**Gallery of Cluster Galaxies**
Use the Simple Image Access (SIA) service to retrieve images from the Gemini GOGREEN program.

## Intended for

- Training
- Education
- Research

## Science cases

- Intro
- Galactic
- Extragalactic
- Time Domain

## Datasets

- Images
- Catalogs
- Spectra

# Introduction to SPARCL

# SPARCL

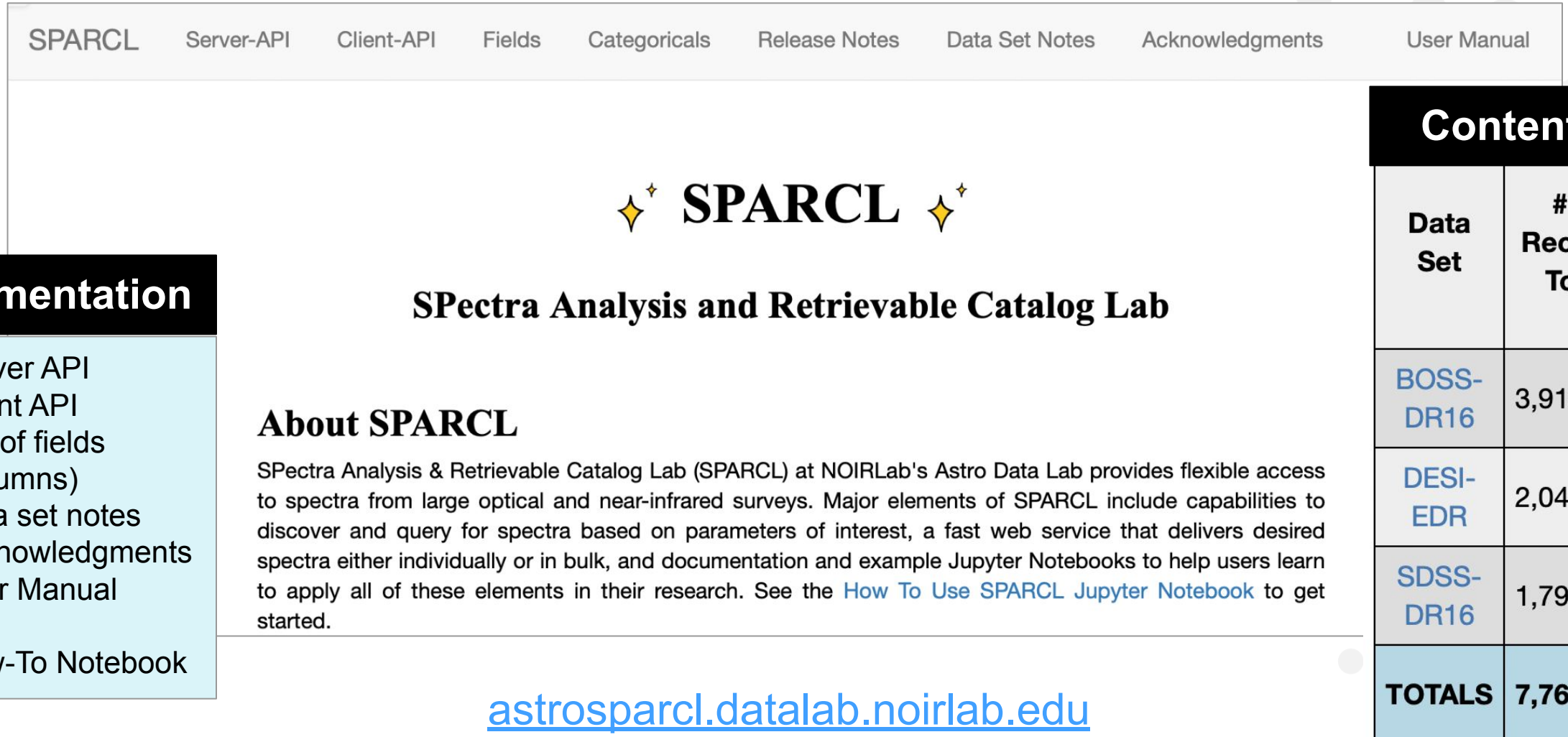
SPectra Analysis & Retrievable Catalog Lab

astrosparcl.datalab.noirlab.edu

- Spectroscopic database for large surveys/datasets
- Data Discovery
- Data Access/Retrieval
- Server can work with different clients
- Compatible with Astro Data Lab (works in JupyterLab)
- Public version: SDSS DR16 (SDSS, (e)BOSS), DESI EDR

- Future goals:
  - Add other spectroscopic datasets (streamlining the ingest process)
  - Add more advanced functionality (e.g., aligning spectra)
  - Add an *authentication* feature (e.g., proprietary access for DESI collaborators)

# Website




## Documentation

- Server API
- Client API
- List of fields (columns)
- Data set notes
- Acknowledgments
- User Manual

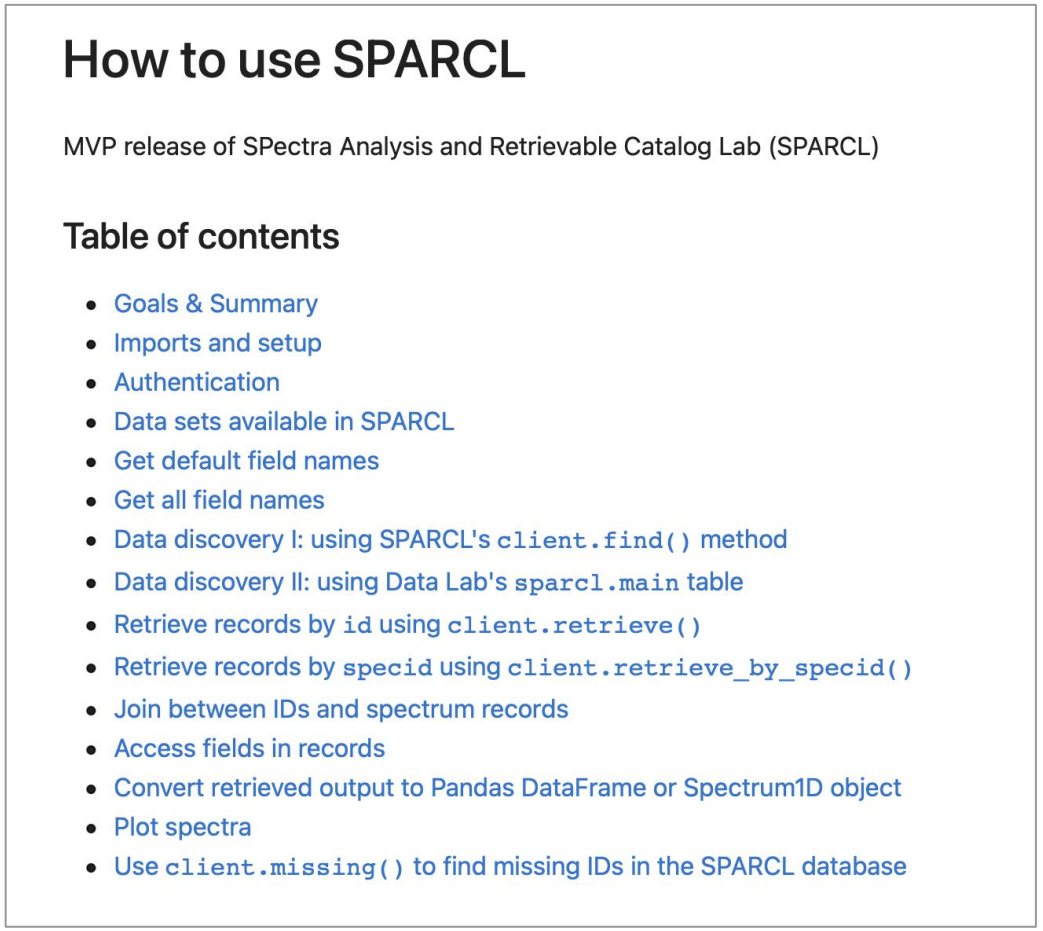
\+ How-To Notebook

## Contents

| Data Set | # of Records Total |
|---|---|
| BOSS-DR16 | 3,918,000 |
| DESI-EDR | 2,044,588 |
| SDSS-DR16 | 1,798,901 |
| **TOTALS** | **7,761,489** |

astrosparcl.datalab.noirlab.edu

# Notebook: How to use SPARCL



## How to use SPARCL

MVP release of SPectra Analysis and Retrievable Catalog Lab (SPARCL)

### Table of contents



- Public version for SDSS DR16 & now DESI EDR
- Fast data discovery and access to 1D spectra (retrieval for up to 20k spectra per call)

Full notebook available (Jacques et al.):
github.com/astro-datalab/notebooks-latest/blob/master/04_HowTos/SPARCL/How_to_use_SPARCL.ipynb

# Contact

datalab@noirlab.edu

datalab.noirlab.edu

github.com/astro-datalab

@DataLabAstro

#datalab-tutorial

Thank you!

Hacking together on great science with Data Lab!

# Resources

Data Lab website: https://datalab.noirlab.edu

SPARCL website: https://astrosparcl.datalab.noirlab.edu

Register account: https://datalab.noirlab.edu/account/register.html

User manual: https://datalab.noirlab.edu/docs/manual

Helpdesk: https://datalab.noirlab.edu/help

Notebook server: https://cloud.datalab.noirlab.edu

Notebook collection: https://github.com/astro-datalab/notebooks-latest

Notebook gallery: [LINK]

Tutorial materials: https://github.com/astro-datalab/Tutorial-ADASS-2023

How to clone them:

```
git clone https://github.com/astro-datalab/Tutorial-ADASS-2023.git
```

# Backup slides