

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

*Astro Data Lab Team <[datalab@noirlab.edu](mailto:datalab@noirlab.edu)>*

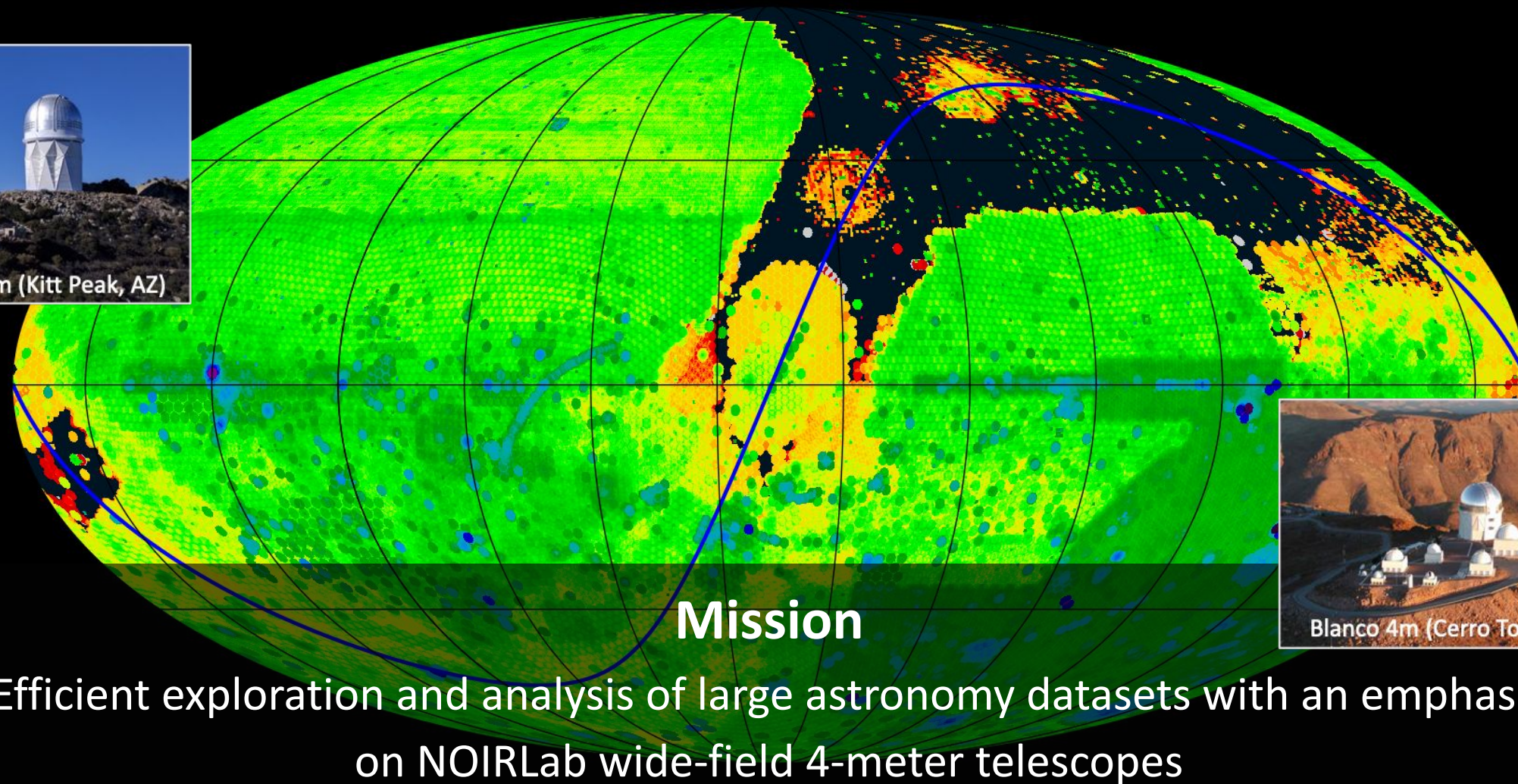
NSF's NOIRLab

Tutorial at ADASS XXXIII, November 2023, Tucson/AZ

# Introduction to the Astro Data Lab science platform



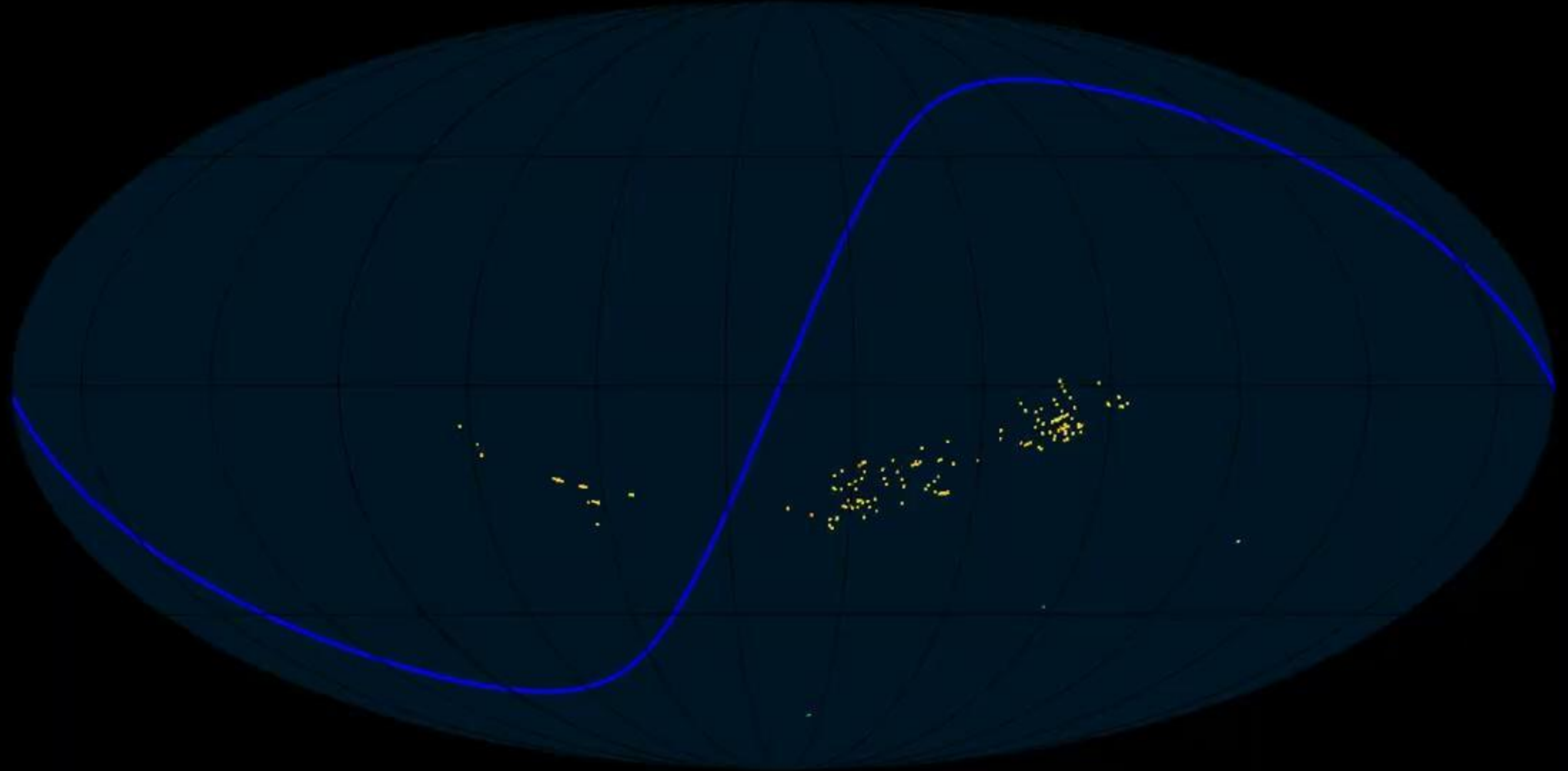
# The Astro Data Lab



## Mission

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

2004.6



Total exposure time on NOIRLab 4-m class telescopes





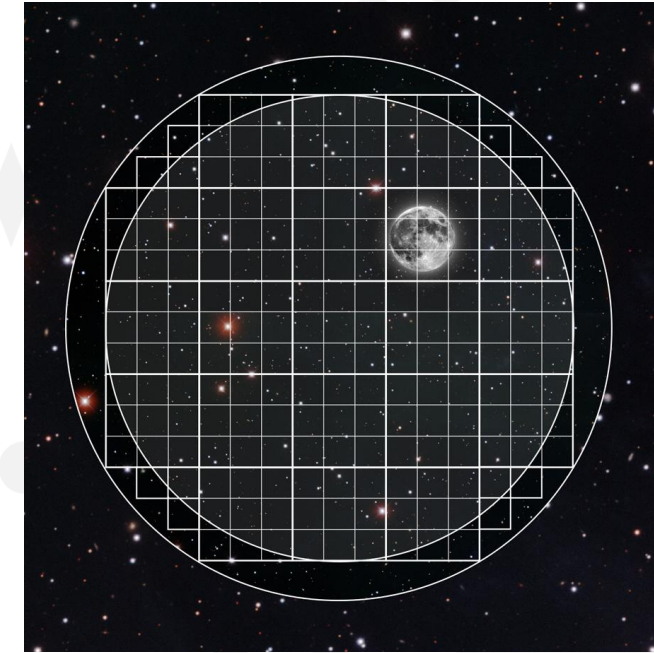
# Wide-field cameras & data avalanche



**Dark Energy Camera**  
570 Mpix



**Rubin Camera**  
3200 Mpix



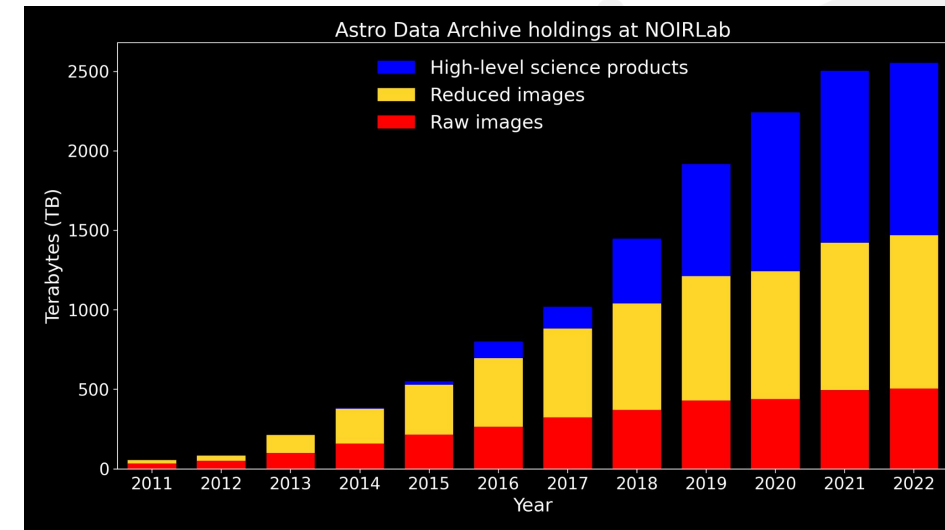
**Rubin Camera**  
40x Moon area

## 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
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
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

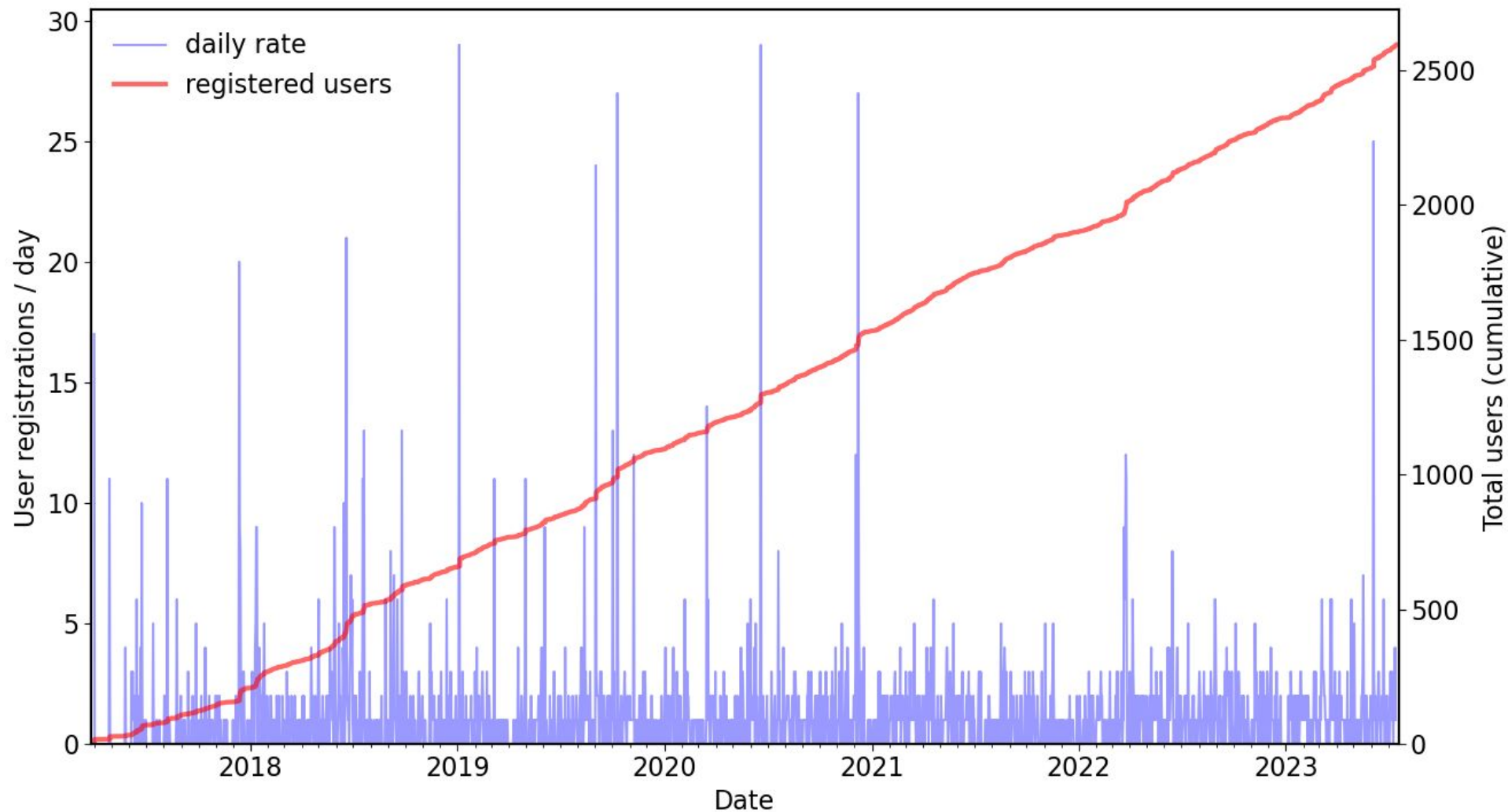
**NOIRLab  
telescopes**

**external**

**simulated**



# Growing user community





# 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

astro-datalab / notebooks-latest Public

<> Code Issues 8 Pull requests Discussions Actions Projects Wiki Security

master 11 branches 0 tags

Go to file Add file Code

rnikutta Merge pull request #142 from astro-datalab/keywords 5595b69 3 days ago 528 commits		
01_GettingStartedWithDataLab	Replace image cutout example with a globular cluster. Clean up and s...	4 months ago
02_DataAccessOverview	- noao to noirlab NB update batch 1 (both ipynb and html files)	10 months ago
03_ScienceExamples	Delete temp	28 days ago
04_HowTos	Replace gaia_edr3 with gaia_dr3	2 months ago
05_Contrib	Update antares notebooks to use the updated kernel	last month
06_EPO	Changed cell order in BlackHole NB	2 months ago
tests	- Exclude e-TeenAstronomyCafe NBs from default testing.	16 months ago
.gitignore	Set lineWidth parameter to 3, on lightcurve_slider.py in order to solv...	12 months ago
CONTRIBUTING	Update README.txt and CONTRIBUTING with new 'astro-datalab' G...	7 months ago
DataLabNotebookTemplate.ipynb	noaodatalab --> astro-datalab	7 months ago
LICENSE	- noao to noirlab NB update batch 1 (both ipynb and html files)	10 months ago

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

Data Lab 1.2.0 documentation

## Welcome to the Astro Data Lab documentation

Contents:

- 1. Using Data Lab
  - 1.1. Introduction
  - 1.2. Science examples
  - 1.3. Web Interfaces
  - 1.4. Data Access Interfaces
  - 1.5. Client Interfaces
  - 1.6. Service Interfaces
  - 1.7. Command-Line Tools
  - 1.8. Install Datalab
  - 1.9. Jupyter Notebooks & JupyterLab
  - 1.10. Jupyter Notebooks Classic
  - 1.11. Upload External Data
  - 1.12. Compute Processing
  - 1.13. Guidance on constructing queries
  - 1.14. SQL Gotchas
  - 1.15. Helpful Resources
  - 1.16. Known Issues
- 2. FAQs
- 3. Developer Guides
  - 3.1. Exception Handling in Data Lab Services
  - 3.2. Documenting Python APIs with Docstrings
  - 3.3. Python Performance Profiling
  - 3.4. Data Lab Python Style Guide
  - 3.5. Python Unit Testing
  - 3.6. ReStructuredText Sample
  - 3.7. ReStructuredText Style Guide
  - 3.8. Page Title Example
- 4. Appendices
  - 4.1. Example Queries
  - 4.2. Notebooks
  - 4.3. Python 3 vs Python 2
  - 4.4. Glossary
- 5. Data Publication docs
  - 5.1. Data Provider Documentation
  - 5.2. Data Format Definitions
  - 5.3. Data Provider Web Template

## Indices and tables

- Index
- Search Page

Data Lab 1.2.0 documentation

**User Manual & API docs**

Welcome to the Data Lab User Forum

All Activity Questions Unanswered Tags Ask a Question

## Astro Data Lab User Forum

Search answers for all your queries One destination for all your queries Get answers from the experts

### Recent questions and answers

Ask a question:

Cannot Upload Local .csv Table to MyDB through Jupyter

answered Jun 30 by anonymous 22 views

mydb\_import mydb

0 votes 2 answers

383 questions

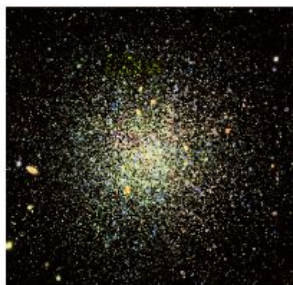
**Helpdesk & FAQ**



# Discovery Ready Science Notebooks

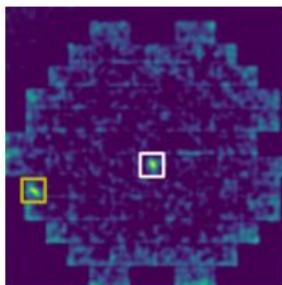


## Astro Data Lab Notebook Gallery



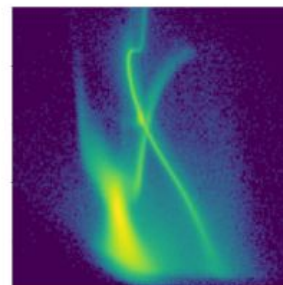
### Getting Started with Data Lab

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



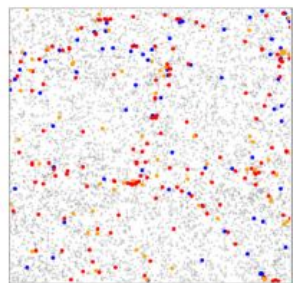
### Dwarf Galaxies in the SMASH survey

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



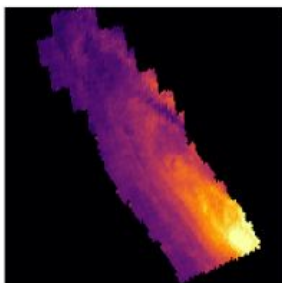
### Exploring SMASH DR2

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.



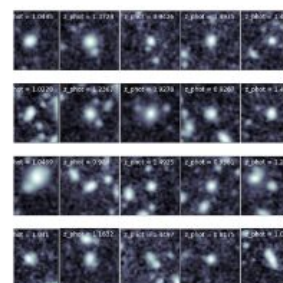
### Large-Scale Structure of the Universe

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







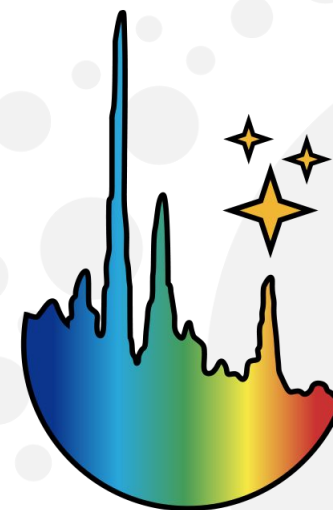
# SPARCL

Spectra Analysis & Retrievable Catalog Lab

[astrosparcl.datalab.noirlab.edu](http://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)



**SPARCL**



# Website

[SPARCL](#)[Server-API](#)[Client-API](#)[Fields](#)[Categoricals](#)[Release Notes](#)[Data Set Notes](#)[Acknowledgments](#)[User Manual](#)

## Documentation

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

+ How-To Notebook



## SPECTRA ANALYSIS AND RETRIEVABLE CATALOG LAB

### About SPARCL

SPECTRA ANALYSIS & RETRIEVABLE CATALOG LAB (SPARCL) at NOIRLab's Astro Data Lab provides flexible access to spectra from large optical and near-infrared surveys. Major elements of SPARCL include capabilities to discover and query for spectra based on parameters of interest, a fast web service that delivers desired spectra either individually or in bulk, and documentation and example Jupyter Notebooks to help users learn to apply all of these elements in their research. See the [How To Use SPARCL Jupyter Notebook](#) to get started.

[astroparcl.datalab.noirlab.edu](https://astroparcl.datalab.noirlab.edu)

## 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



# Notebook: How to use SPARCL



## How to use SPARCL

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

### Table of contents

- [Goals & Summary](#)
- [Imports and setup](#)
- [Authentication](#)
- [Data sets available in SPARCL](#)
- [Get default field names](#)
- [Get all field names](#)
- [Data discovery I: using SPARCL's `client.find\(\)` method](#)
- [Data discovery II: using Data Lab's `sparcl.main` table](#)
- [Retrieve records by id using `client.retrieve\(\)`](#)
- [Retrieve records by specid using `client.retrieve\_by\_specid\(\)`](#)
- [Join between IDs and spectrum records](#)
- [Access fields in records](#)
- [Convert retrieved output to Pandas DataFrame or Spectrum1D object](#)
- [Plot spectra](#)
- [Use `client.missing\(\)` to find missing IDs in the SPARCL database](#)

- 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](https://github.com/astro-datalab/notebooks-latest/blob/master/04_HowTos/SPARCL/How_to_use_SPARCL.ipynb)



# Contact



[datalab@noirlab.edu](mailto:datalab@noirlab.edu)



[datalab.noirlab.edu](http://datalab.noirlab.edu)



[github.com/astro-datalab](https://github.com/astro-datalab)



[@DataLabAstro](https://twitter.com/DataLabAstro)



[#datalab-tutorial](https://twitter.com/DataLabAstro)

## Thank you!



Hacking together on great science with Data Lab!



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

SPARCL website: <https://astroparcl.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

