

Exploring big data efficiently with the Astro Data Lab science platform

Astro Data Lab Team <datalab@noirlab.edu>

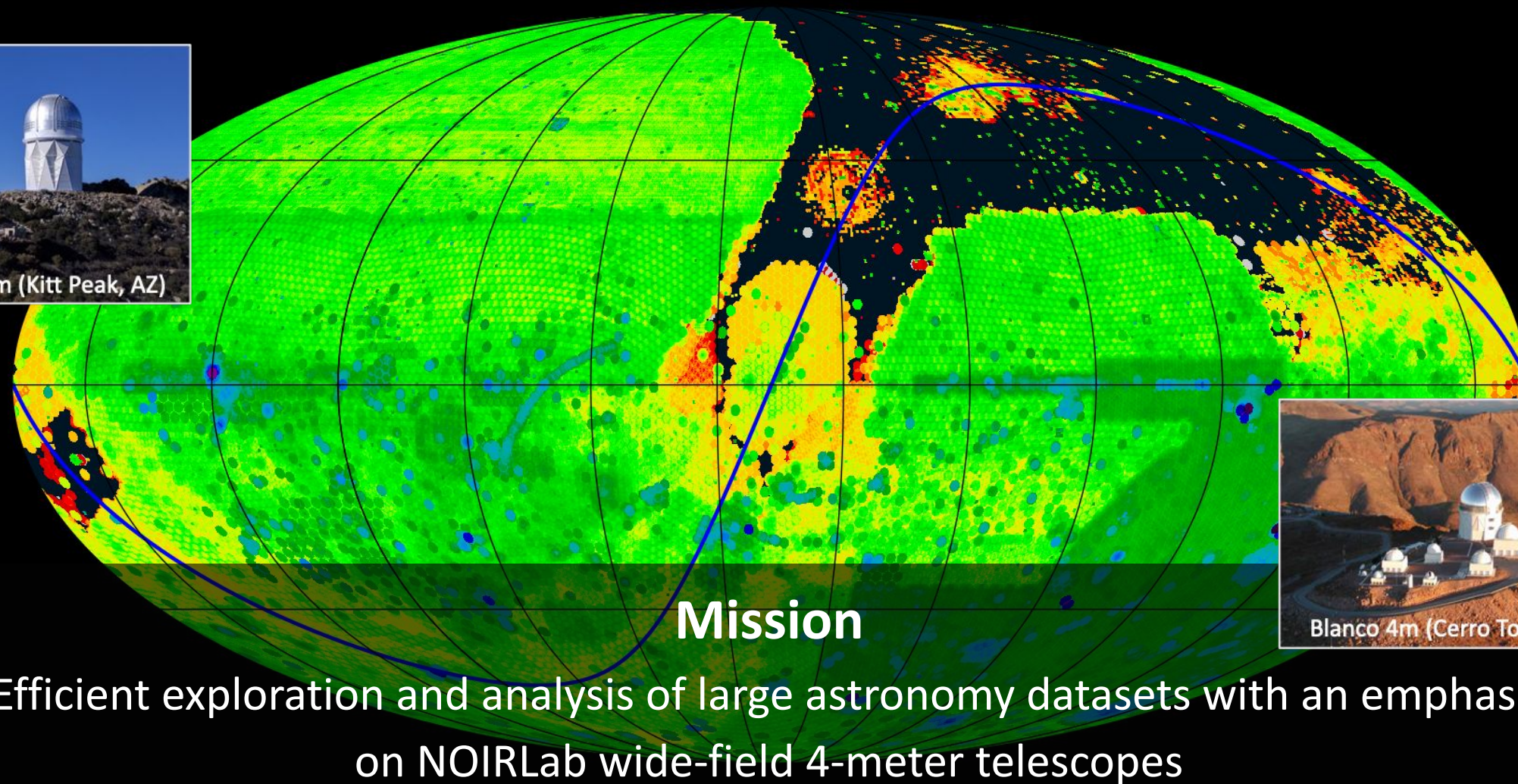
NSF's NOIRLab

Tutorial at ADASS XXXIII, November 2023, Tucson/AZ

Introduction to the Astro Data Lab science platform



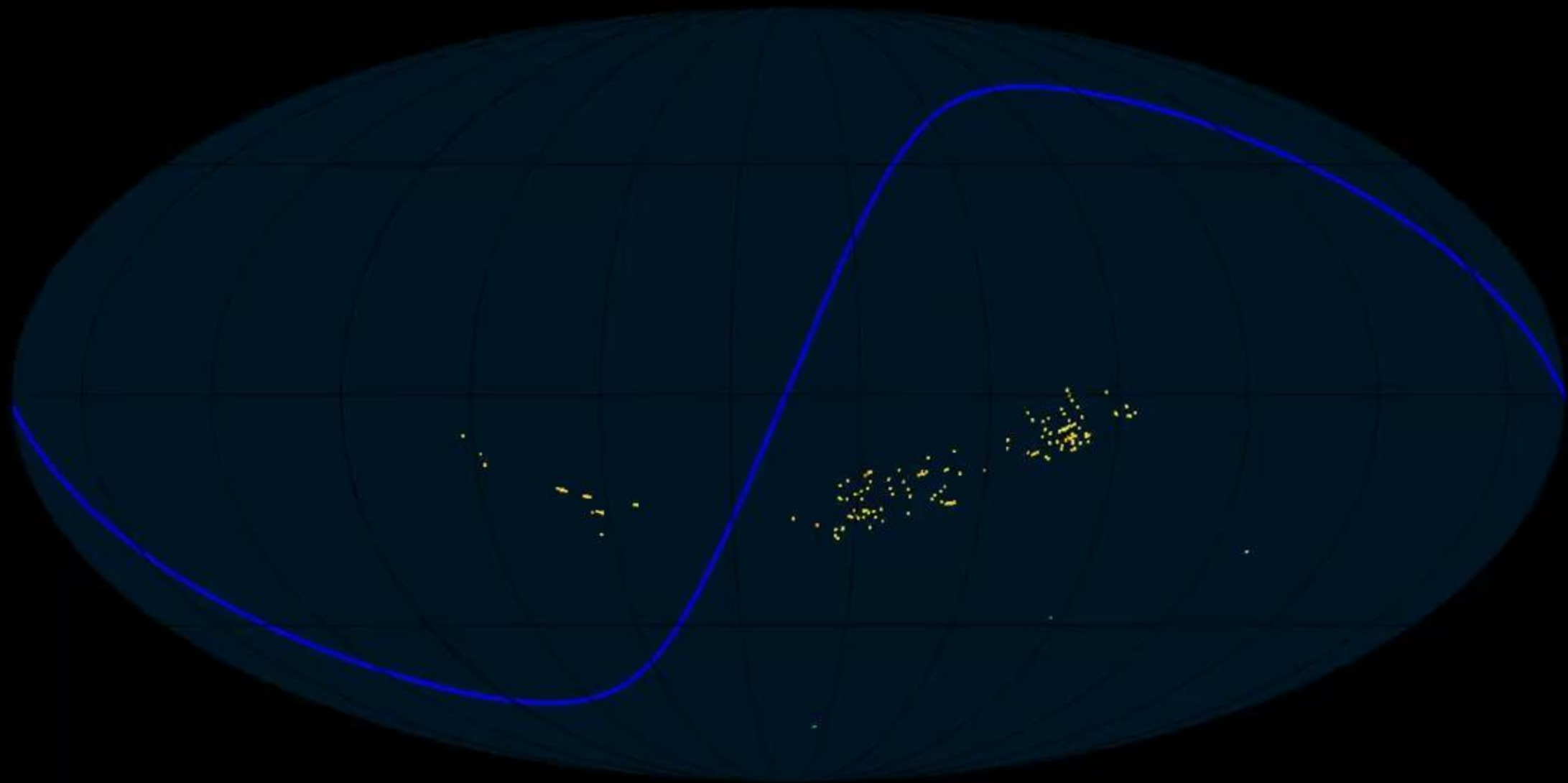
The Astro Data Lab



Mission

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

2004.6



Total exposure time on NOIRLab 4-m class telescopes



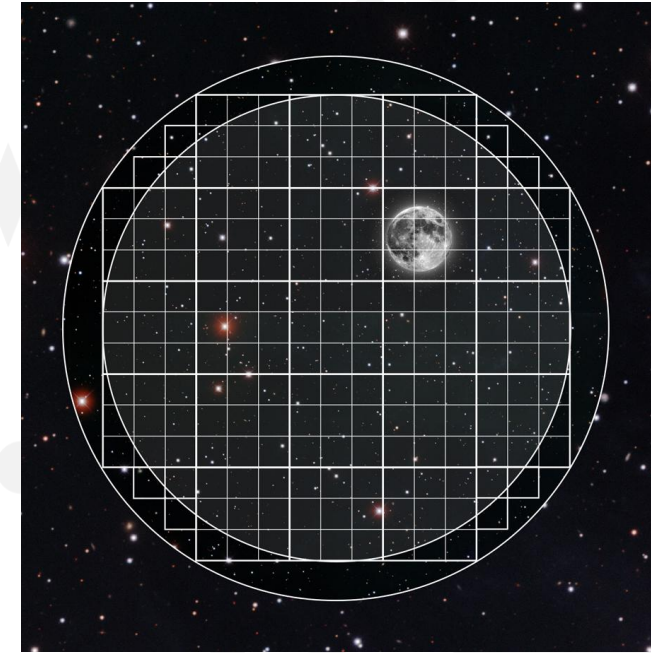
Wide-field cameras & data avalanche



Dark Energy Camera
570 Mpix



Rubin Camera
3200 Mpix



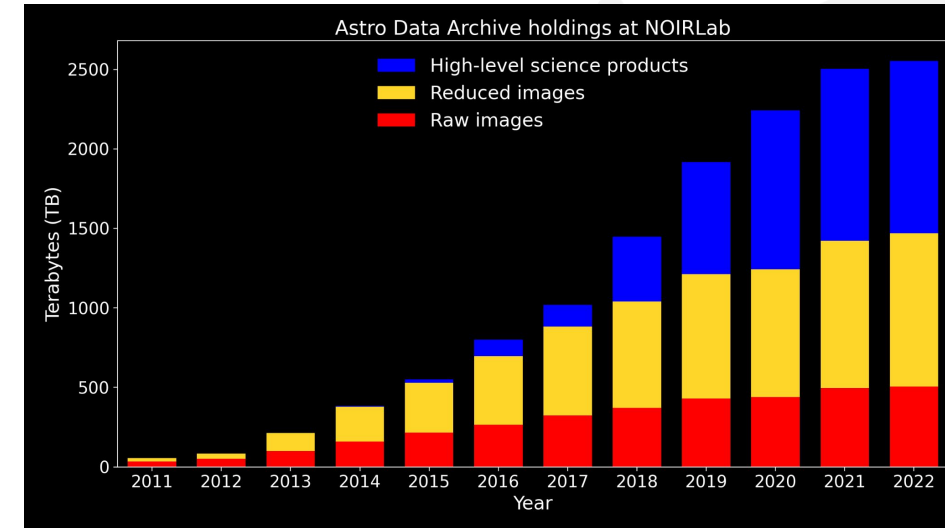
Rubin Camera
40x Moon area

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
DECaLS DR 3 through 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

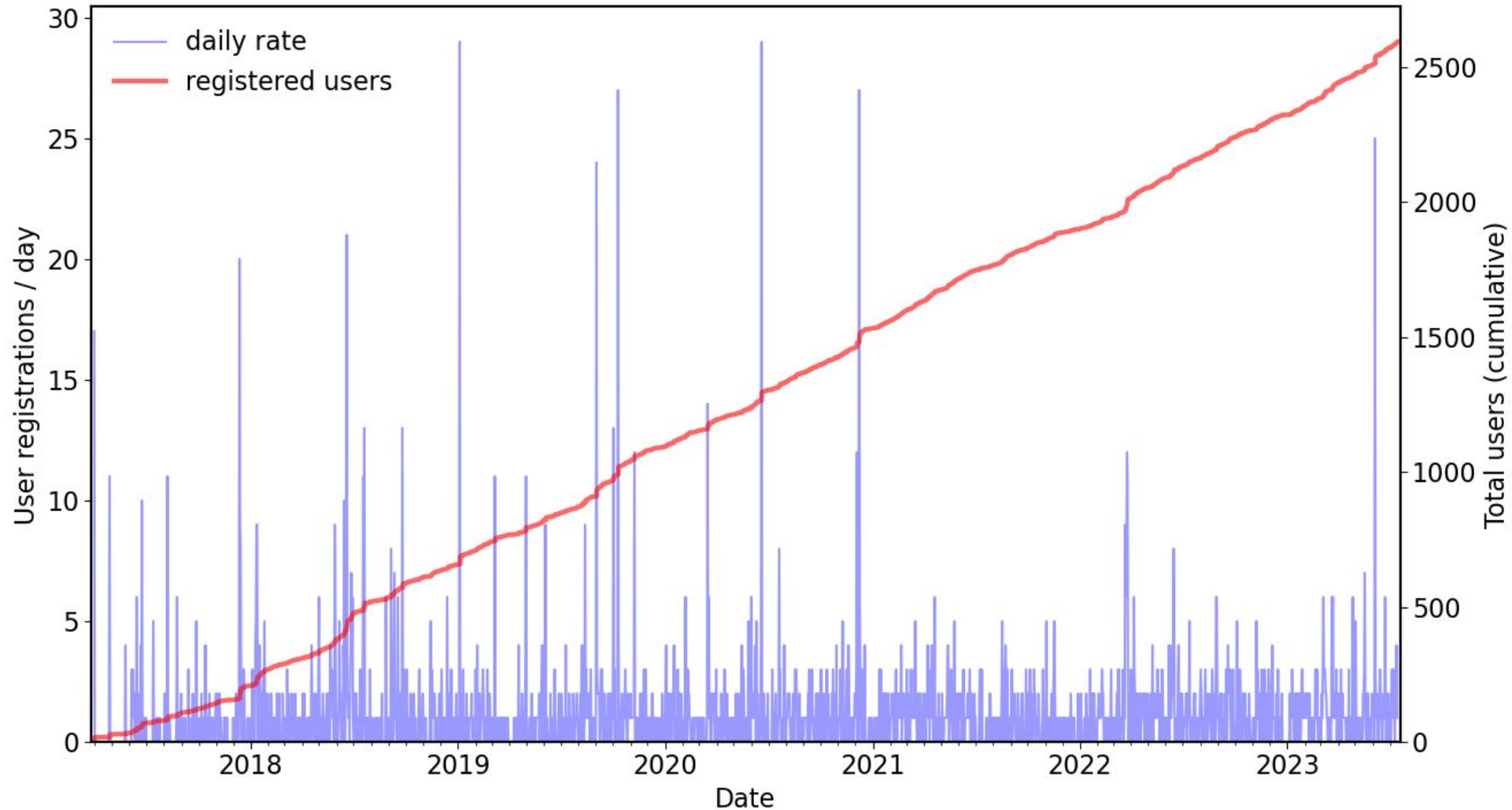
**NOIRLab
telescopes**

external

simulated



Growing user community



As of June 14, 2023

Data services & tools

Action	Modes	Notes
Authentication	web, client/API, cmdline	log on to access user-specific functions
Sky exploration	Aladdin Lite	
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	5505b60 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...

File Edit View Tools Window Help Data Lab documentation — NOAO Data Lab

NOAO Data Lab 1.1.0 documentation » next | index

Welcome to the NOAO Data Lab documentation

Contents:

- 1. Using the NOAO 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 Data Lab
 - 1.9. Jupyter Notebooks
 - 1.10. Compute Processing
 - 1.11. Guidance on constructing queries
 - 1.12. SQL gotchas
 - 1.13. Helpful Resources
 - 1.14. Known Issues
- 2. FAQs
- 3. Appendices
 - 3.1. Example Queries
 - 3.2. Notebooks
- 4. Data Publication docs
 - 4.1. Data Provider Documentation
 - 4.2. Data Format Definitions
 - 4.3. Data Provider Web Template

Table Of Contents

Welcome to the NOAO Data Lab documentation

Indices and tables

Next topic

1. Using the NOAO Data Lab

This Page

Show Source

Quick search

Go

NOAO Data Lab 1.1.0 documentation » next | index

© Copyright 2016-2017. NOAO Data Lab <datalab@noao.edu>. Last updated on 2018-3an-07.

User Manual & API docs

File Edit View Tools Window Help Data Lab Help Desk

Data Lab Home NOAO Data Lab User Forum

Forum home All Activity Questions Unanswered Tags Categories Users Ask a Question

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:

Unable to login to Datalab

answered 5 days ago by kolsen (300 points) | 0 views

0 votes

1 answer

Search

68 questions

Helpdesk & FAQ



Discovery Ready Science Notebooks

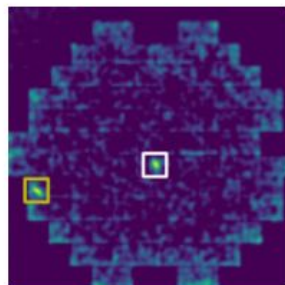


Astro Data Lab Notebook Gallery



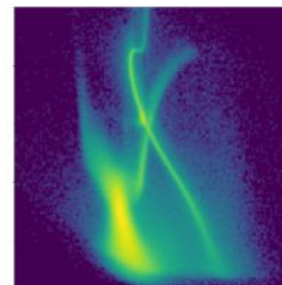
Getting Started with Data Lab

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



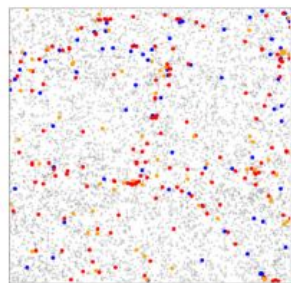
Dwarf Galaxies in the SMASH survey

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



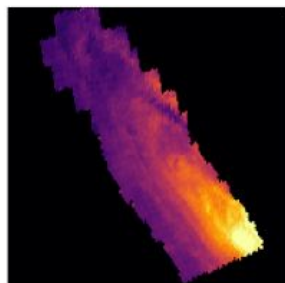
Exploring SMASH DR2

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.



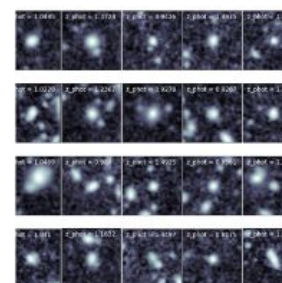
Large-Scale Structure of the Universe

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





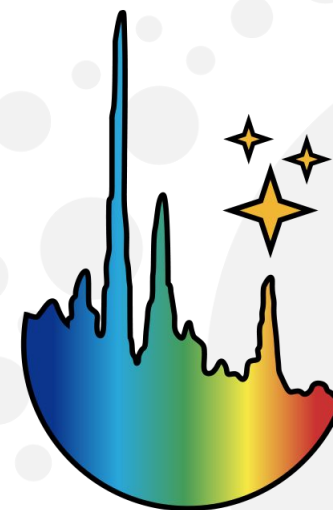
SPARCL

Spectra Analysis & Retrievable Catalog Lab

astrosparcl.datalab.noirlab.edu



- 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](#)

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

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



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



Contact



datalab@noirlab.edu



datalab.noirlab.edu



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

