Exploring big data efficiently with the **Astro Data Lab** science platform and **SPARCL** spectral database

Brian Merino,
on behalf of the Astro Data Lab and SPARCL teams

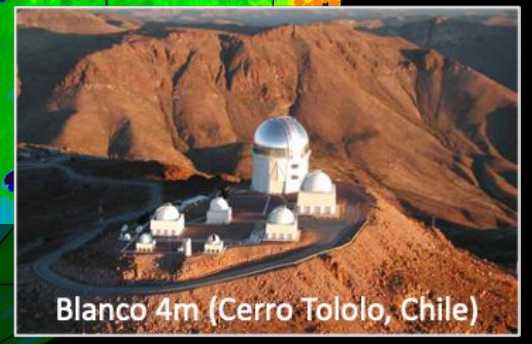
ADASSx, Tucson AZ, August 2025

The Astro Data Lab & SPARCL Teams

Robert **Nikutta**, Data Lab Project Scientist [online]
Stephanie **Juneau**, SPARCL Project Scientist [in person]

David **Herrera**, Data Analyst [in person]
Alice **Jacques**, Data Analyst [in person]
Sean **McManus**, SPARCL Software Manager
Brian **Merino**, Data Analyst [in person]
Chadd **Myers**, Software Engineer [online]
Peter **Peterson**, Software Engineer
Benjamin **Weaver**, Scientist [in person]
Rohin **Sant**, TIMESTEP Intern [in person]
Joel **Shernicoff**, TIMESTEP Intern [online]

The Astro Data Lab



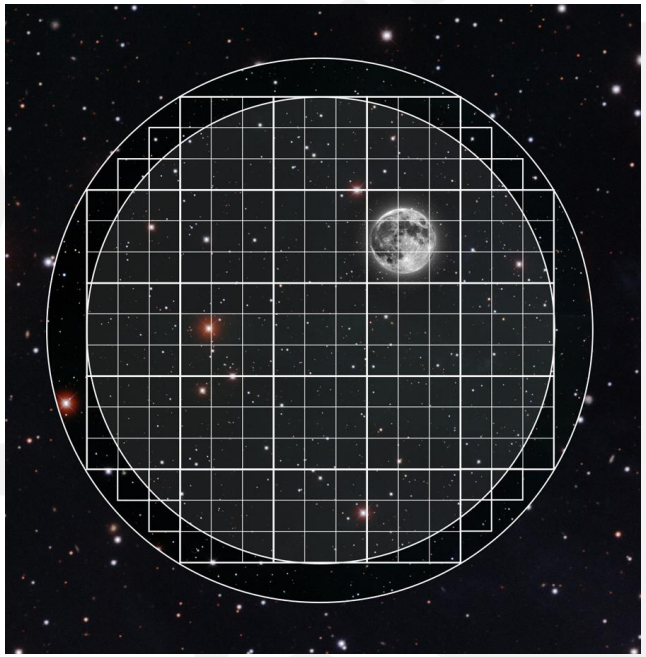
Mission

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

Wide-field cameras & data avalanche

Dark Energy Camera
570 Mpix
(2012-present)

Rubin Camera
3200 Mpix
(2025-present ; 20 TB/night)

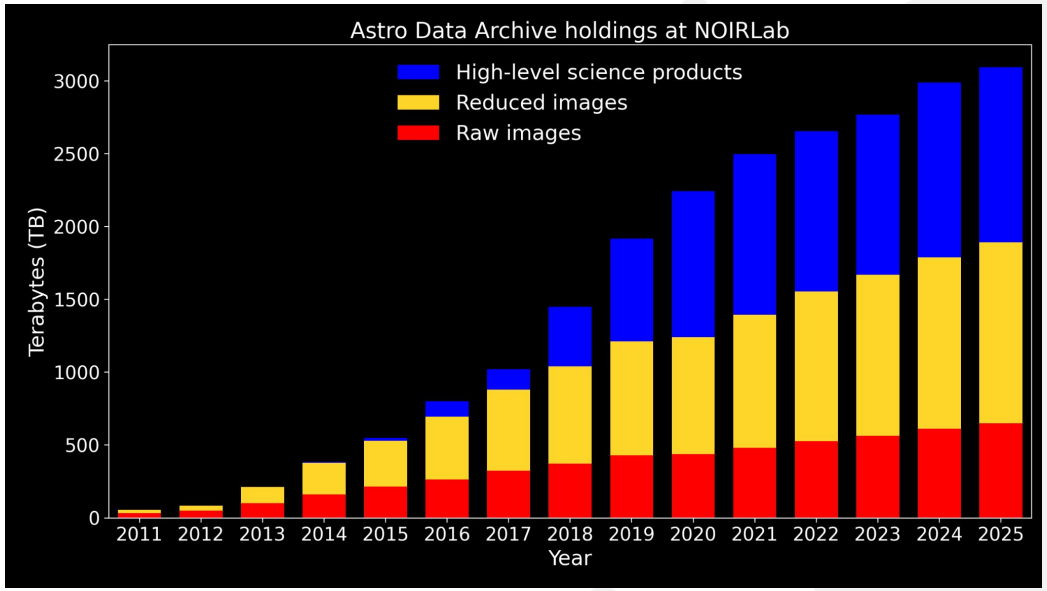


Rubin Camera
40x Moon area

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

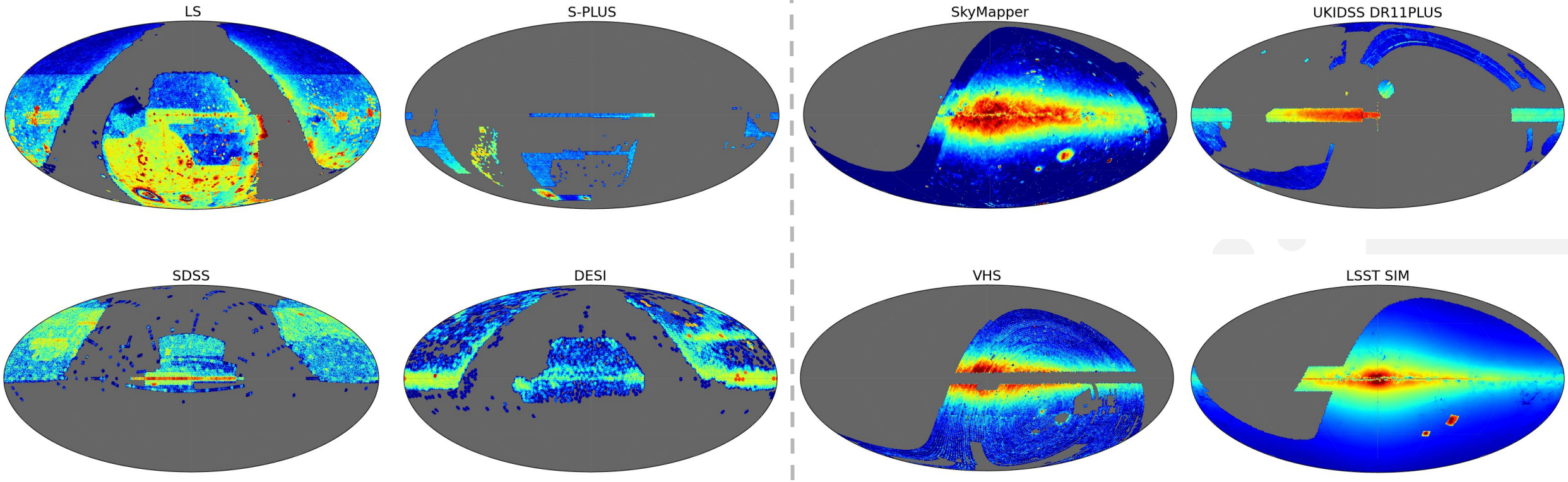
Dataset	Number of objects or measurements	Notes
DES DR 1 & 2	691M objects in DR 2	photometry
DESI Legacy Surveys DR 8 - 10	3.14B objects in DR 10	photometry, targeting for DESI
NOIRLab Source Catalog DR 1 & 2	3.9B objects, 68B measurements in DR 2	homogen. reduction across 3 cameras
SMASH DR 1 & 2, DELVE DR 1 & 2	2.5B objects in DELVE DR 2	photometry
DESI EDR & DR1, SPARCL (metadata)	23M spectra in DESI DR1	spectroscopy
Gaia DR 1, 2, EDR3, 3	1.8B objects in DR 3	astrometry
AllWISE, unWISE, CatWISE2020	2.2B objects in unWISE	IR
SkyMapper DR 4	724M objects, 15.1B measurements	photometry
S-PLUS DR 2, 4	31.5M objects in DR2	photometry
PHAT v2	117M objects	Andromeda
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

Some datasets hosted at DL

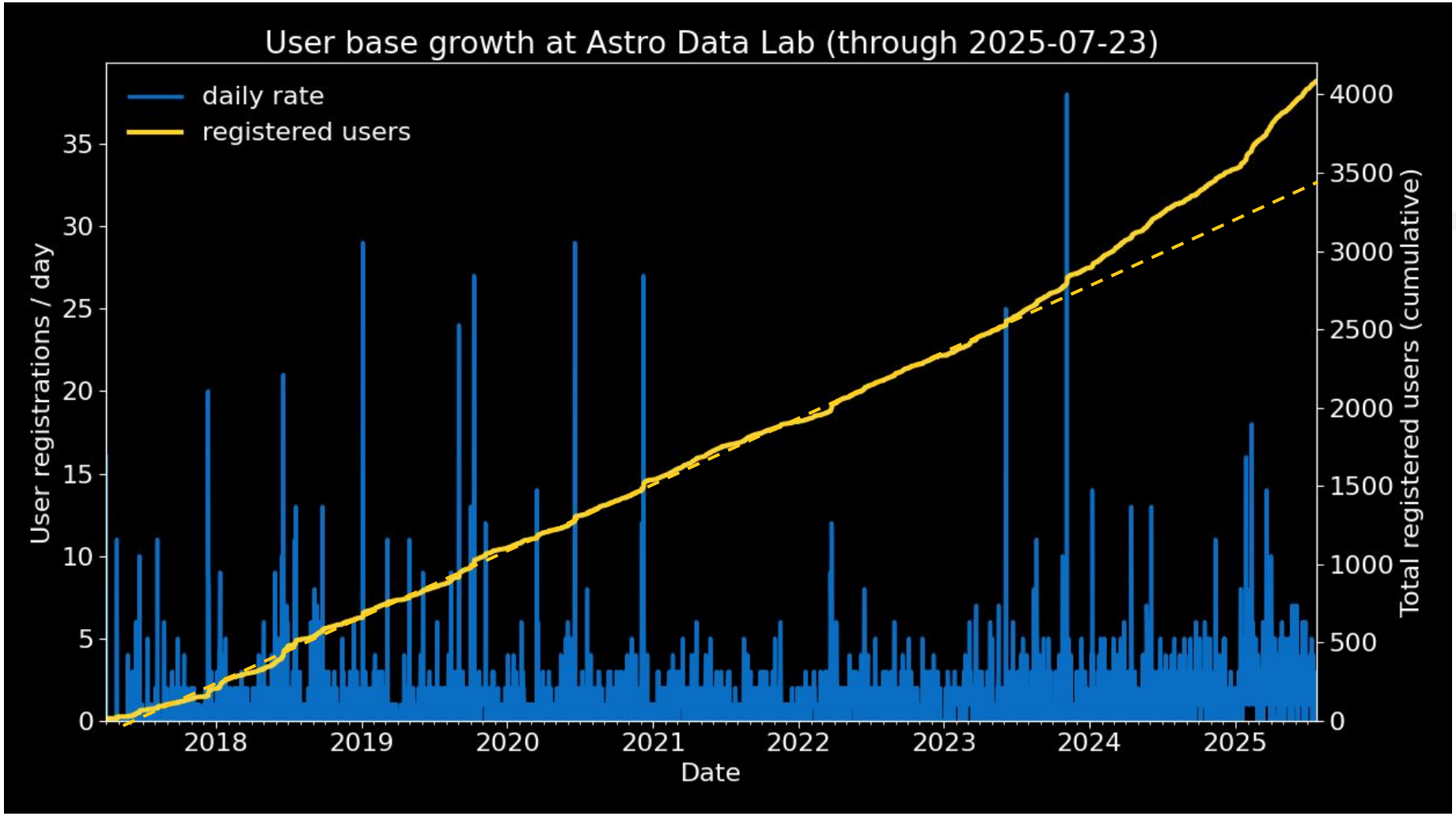


Equatorial coordinates

Galactic coordinates

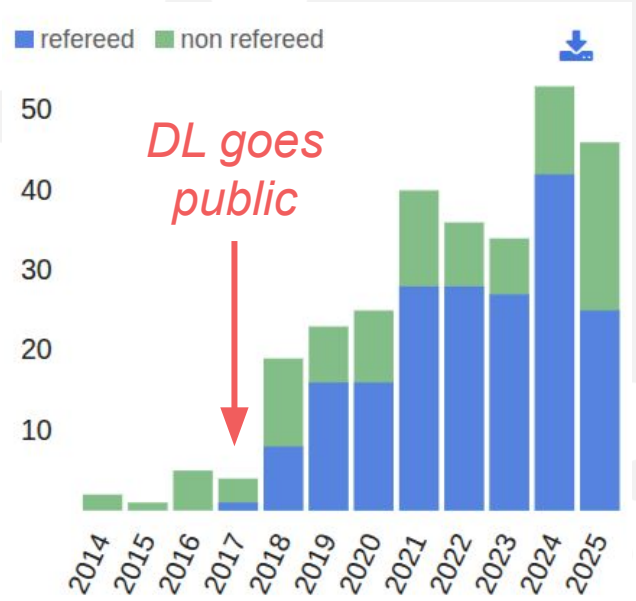
Object number density (per square degree) in Mollweide projection for eight surveys: LS DR10, S-PLUS DR4, SDSS DR17, and DESI DR1 are shown in Equatorial coordinates; SkyMapper DR4, UKIDSS DR11PLUS, VHS DR5, and LSST-SIM DR2 are in Galactic coordinates.

Growing user community



As of July 23, 2025

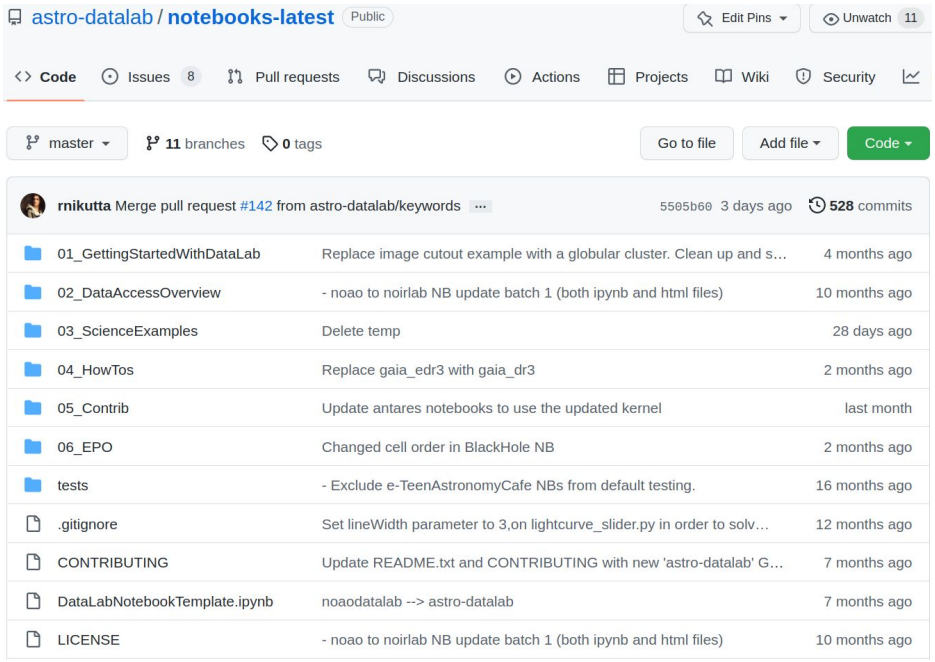
~290 papers so far



Data services & tools

Action	Modes	Notes
Authentication	web, client/API, cmdline	log on to access user-specific functions
Sky exploration	Aladin 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	Web UI being updated; Also hosting pre-xmatched 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



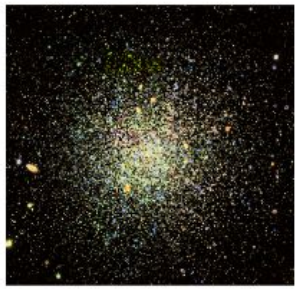
Curated Jupyter notebook collection
Intros, How-Tos, Science Cases, EPO, Contributed…

User Manual & API docs

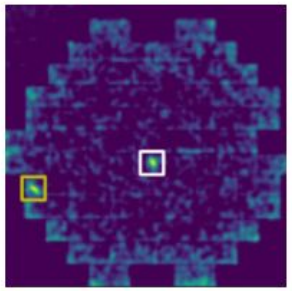
Helpdesk & FAQ

Discovery Ready Science Notebooks

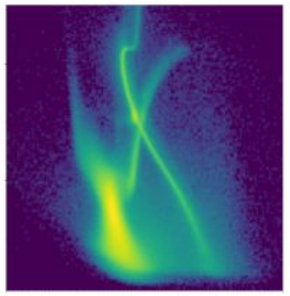
Astro Data Lab Notebook Gallery



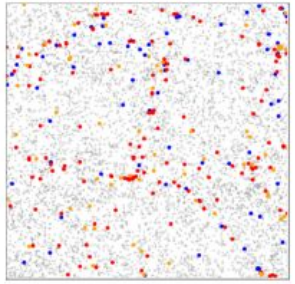
Getting Started with Data Lab
Learn the basics such as importing modules, sending a database query, and using the Simple Image Access (SIA) service to create image cutouts.



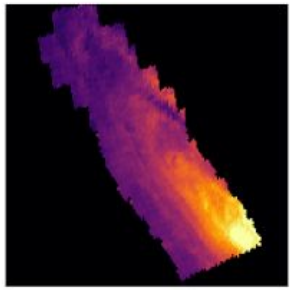
Dwarf Galaxies in the SMASH survey
Discover the ultrafaint Hydra II dwarf galaxy in the SMASH DECam survey based on spatial overdensities of blue stars with a detection algorithm.



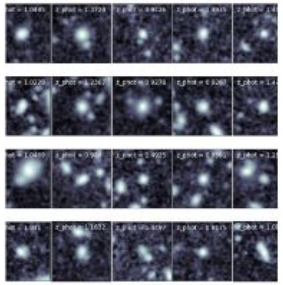
Exploring SMASH DR2
The 480 square degree SMASH DECam survey of the Magellanic Clouds and their periphery contains a wealth of objects, including this capture of the SMC with 47 Tuc in the foreground.



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

Data Lab for Outreach and Education

Teen Astronomy Cafe

- Tucson initiative for high-school students
- Each activity includes: lectures, demos, interactive computer experiments
- Data Labs hosts a collection of notebooks with its NB server

La Serena School for Data Science

- Annual initiative to teach data science to graduate students
- Adaptation and hosting of NBs for a wider audience

Graduate, undergraduate, and high-school astronomy education

- Various universities use the Data Lab science platform for teaching astronomy and big data science
- Data Lab can (and does) host custom kernels, custom data sets, etc.

[Demo of DL website]

datalab.noirlab.edu

Also see ADASSx poster:

“Astro Data Lab’s New Integrated Web Portal”
(Robert Nikutta & Chadd Myers)

Introduction to SPARCL

ADASS Proceedings (arXiv:2401.05576)

SPARCL

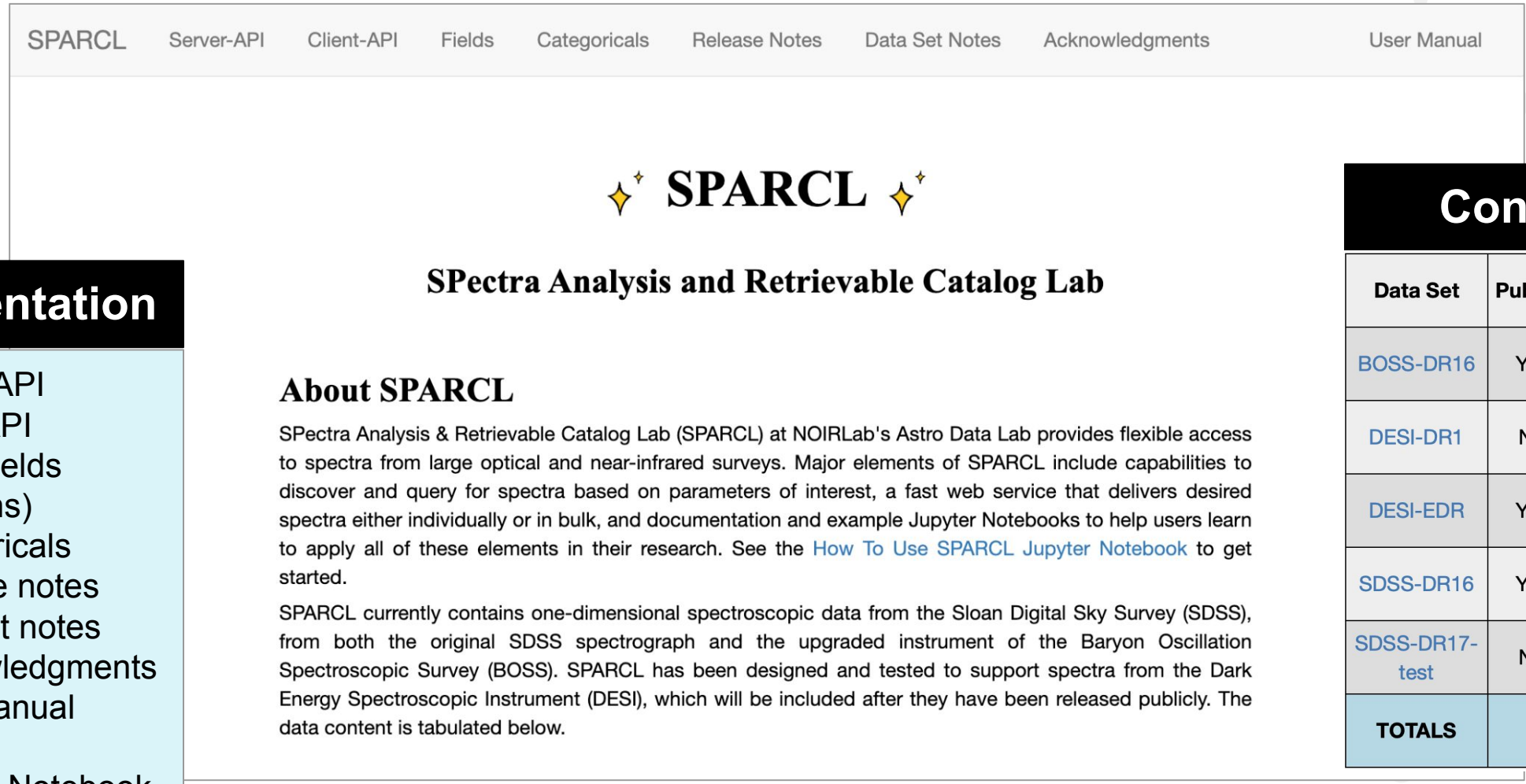
SPectra Analysis & Retrievable Catalog Lab

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)
- Current datasets:
  - **SDSS DR16** (SDSS, BOSS), **DESI EDR**, **DR1** publicly available
  - **SDSS DR17** will replace **DR16** starting on Nov 1st, 2025

- Future goals:
  - Add other spectroscopic datasets (streamlining the ingest process)
  - Add more advanced functionality (e.g., aligning spectra)

Website



Documentation

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

\+ How-To Notebook

Contents

Data Set	Public?	# of Records Total
BOSS-DR16	Yes	3,946,000
DESI-DR1	No	23,060,727
DESI-EDR	Yes	2,044,588
SDSS-DR16	Yes	1,843,200
SDSS-DR17-test	No	1,843,200
TOTALS		**32,737,715**

astrosparcl.datalab.noirlab.edu

***Note:** SDSS/BOSS DR16 will be replaced with SDSS/BOSS DR17

Notebook: How to use SPARCL

```
pip install sparclclient
```

```
from sparcl.client import SparclClient
client = SparclClient()
```

```
found = client.find(outfields=['sparcl_id',
                               'data_release',
                               'ra',
                               'dec',
                               'redshift'],
                    constraints={'spectype': ['GALAXY'],
                                 'redshift': [0.5, 0.9]})
```

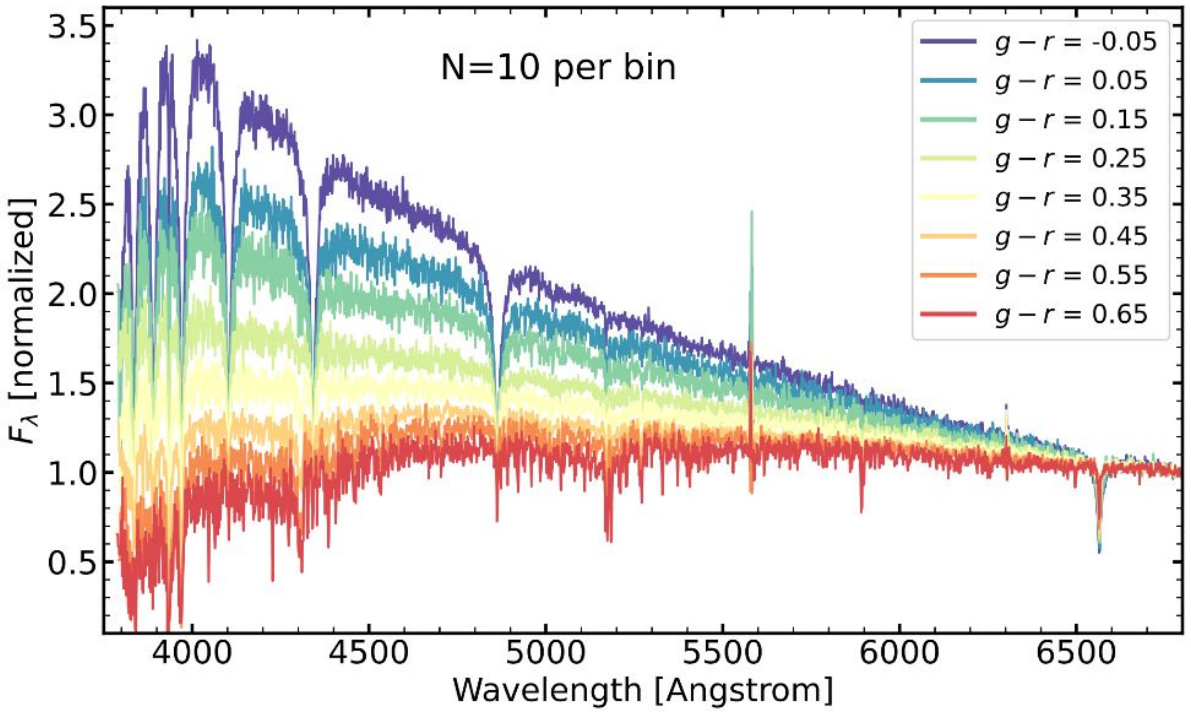
```
results = client.retrieve(uuid_list=found.ids,
                          include=['sparcl_id',
                                   'flux',
                                   'wavelength'])
```

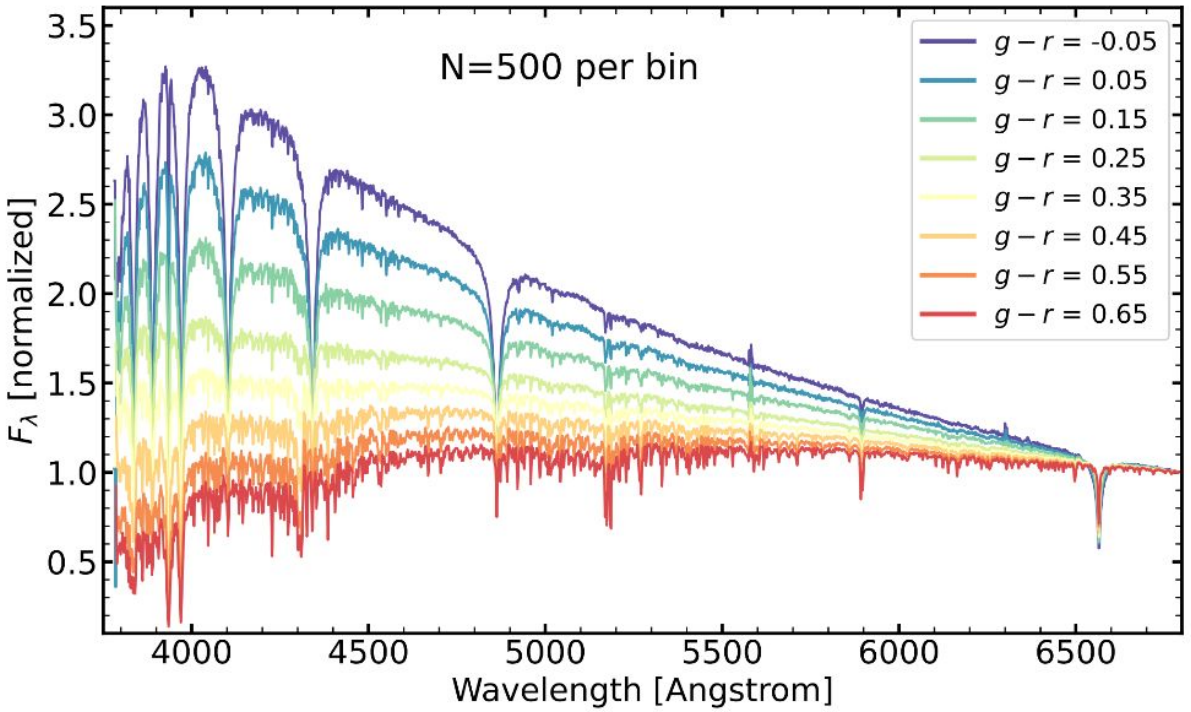
- Public version for SDSS DR16, DESI EDR and DESI DR1
- Fast data discovery and access to 1D spectra (retrieval for up to 20k spectra per call)
- ADASS Proceedings: [arXiv:2401.05576]

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

Example SPARCL use case

1) Using the Astro Data Lab, query “N” stars in eight bins of *g* – *r* color
2) Using SPARCL, retrieve and stack SDSS spectra in each color bin
<u>Results:</u> (left) N=10 takes 5sec for 80 spectra; (right) N=500 takes 90sec for 4000 spectra





(Figure from Juneau et al. 2024; ADASS Proceedings arXiv:2401.05576)

[Example Use Case]

Follow along in your Astro Data Lab account… (optional)

Astro Data Lab Jupyter Notebooks

Currently

- 75 notebooks available:
  - in user accounts
  - on GitHub (github.com/astro-datalab/notebooks-latest/)
- Range of topics, difficulty levels and purposes
  - Introductory tutorials
  - Science cases
  - How-to technical tutorials
  - Education and outreach
- Can be challenging to navigate this expansive collection (especially for new users)

Future plans

- Improve keywords and develop a Notebook Gallery (in progress)

Iyer et al. 2024 (“Pathfinder”)

Notebook Gallery

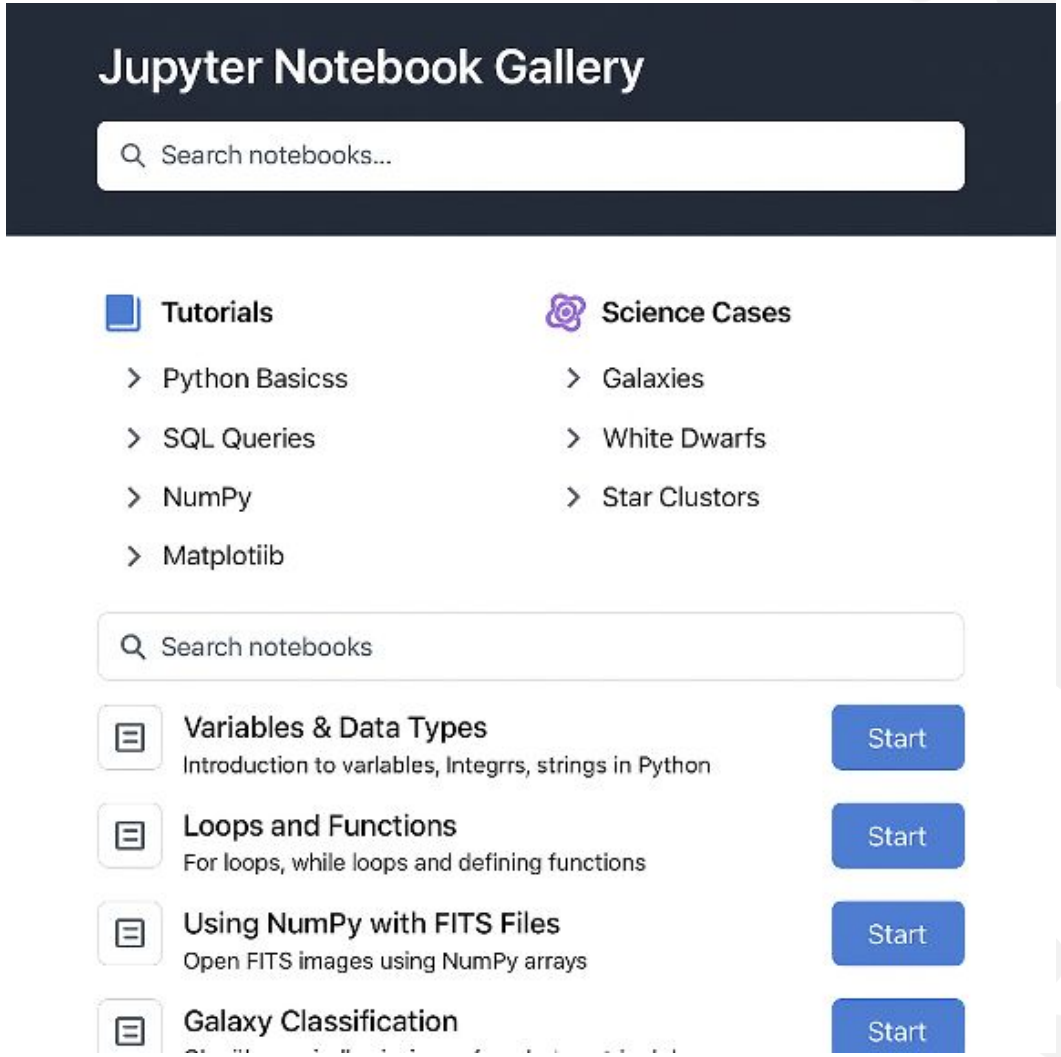
Goals

- Improve notebook discoverability
- Support different audiences (high-school students to researchers)

Future plans

- Add functionality to search, filter and launch notebooks
- Create suggested workflow (navigate to “Previous/Next” or “Similar” notebooks)

example concept
(by TIMESTEP interns R. Sant & J. Shernicoff)



Updated Notebook Keyword System

Changes

- From a single list of keywords to categories
- Topics follow the Unified Astronomy Thesaurus (UAT)
- Added and revised keywords for ease of use

Future plans

- The ability to sort and filter notebooks by keywords
- A semi-automated system to tag notebooks with keywords

Category	Difficulty	Topics	Datasets	Skills	Tools
Tutorial	Introductory	AGN	alwise	Authentication	Ghost
HowTo	Intermediate	LMC	delve_dr2	Convolution	Healpix Map
Science Example	Advanced	Redshift	DESI Spectra	Crossmach	Histogram
Education		SMC	gaia_dr3	Data Reduction	Joint Query
		Solar System	Gemini	Exploratory Analysis	Jupyter
		Star Clusters	gogreen_dr2	Introduction to Machine Learning	Lomb-Scargle
		Transient	SDSS	Supervised Machine Learning	SPARCL
		Variable	smash_dr2	Spectra Stacking	SQL

preliminary (by TIMESTEP interns J. Shernicoff & R. Sant)

CosmicAI Institute

NSF-Simons AI institute for Cosmic Origins led by UT Austin with NOIRLab as a partner

- **Goals:** accelerate astronomy research with transformative AI-assisted workflows while advancing foundational AI
- SPARCL and Astro Data Lab will contribute components of the AI Data Platform
- AI assistant is trained on Astro Data Lab notebooks (and others)
- Example features for an AI-assisted Data Platform:
  - AI-assisted queries for datasets (translation from natural language queries to ADQL)
  - Assistant working in the JupyterLab environment
  - Modular services (SPARCL, Data Lab) with GPU for ML/AI applications

More info (news, mailing list, jobs, etc.): cosmicai.org

Coordinating Spectroscopic Tools

- Full title: Coordinating Spectroscopic Data Reduction and Analysis Tools
- *Anyone interested in spectroscopy software is welcome!* You can participate in more detailed descriptions of our activities, provide feedback and assist in future planning.
- Informal group of spectroscopy software developers and scientists ("SpectroscopyDev").
- This group is an outcome of the meeting "Coordinating the Next Generation of Spectroscopic Processing and Analysis Tools", November 2023.
- Including but not limited to developers of PypeIt, specutils and specreduce, DRAGONS (Gemini), desispec (DESI).
- Benjamin Weaver, `#spectroscopic-tools` on ADASSxTucson Slack.

  13:30 - 15:00 MST. After lunch today.

Contact

datalab@noirlab.edu

datalab.noirlab.edu

github.com/astro-datalab

@AstroDataLab

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/

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

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

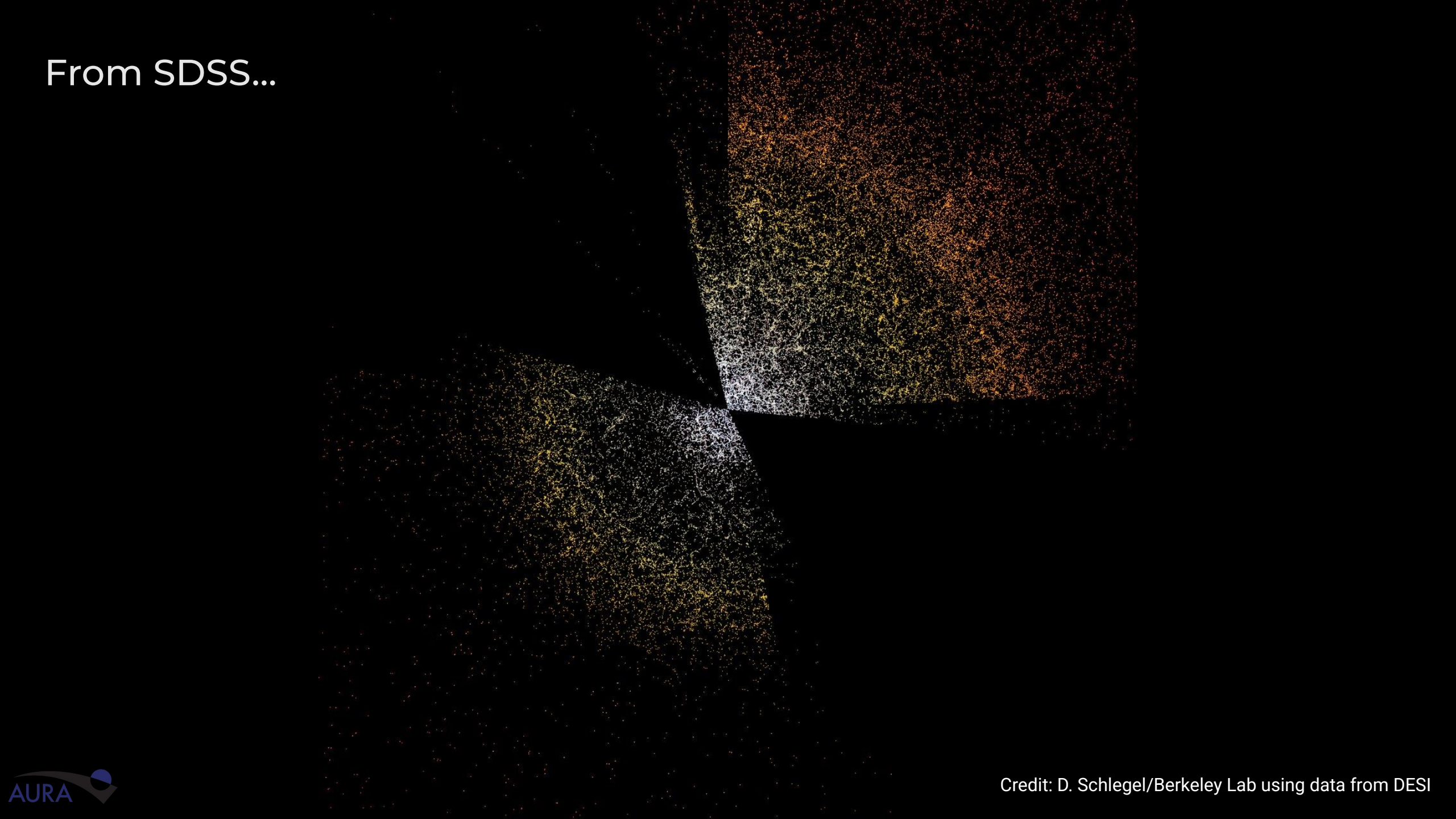
Notebook server: https://datalab.noirlab.edu/notebooks

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

Q&A

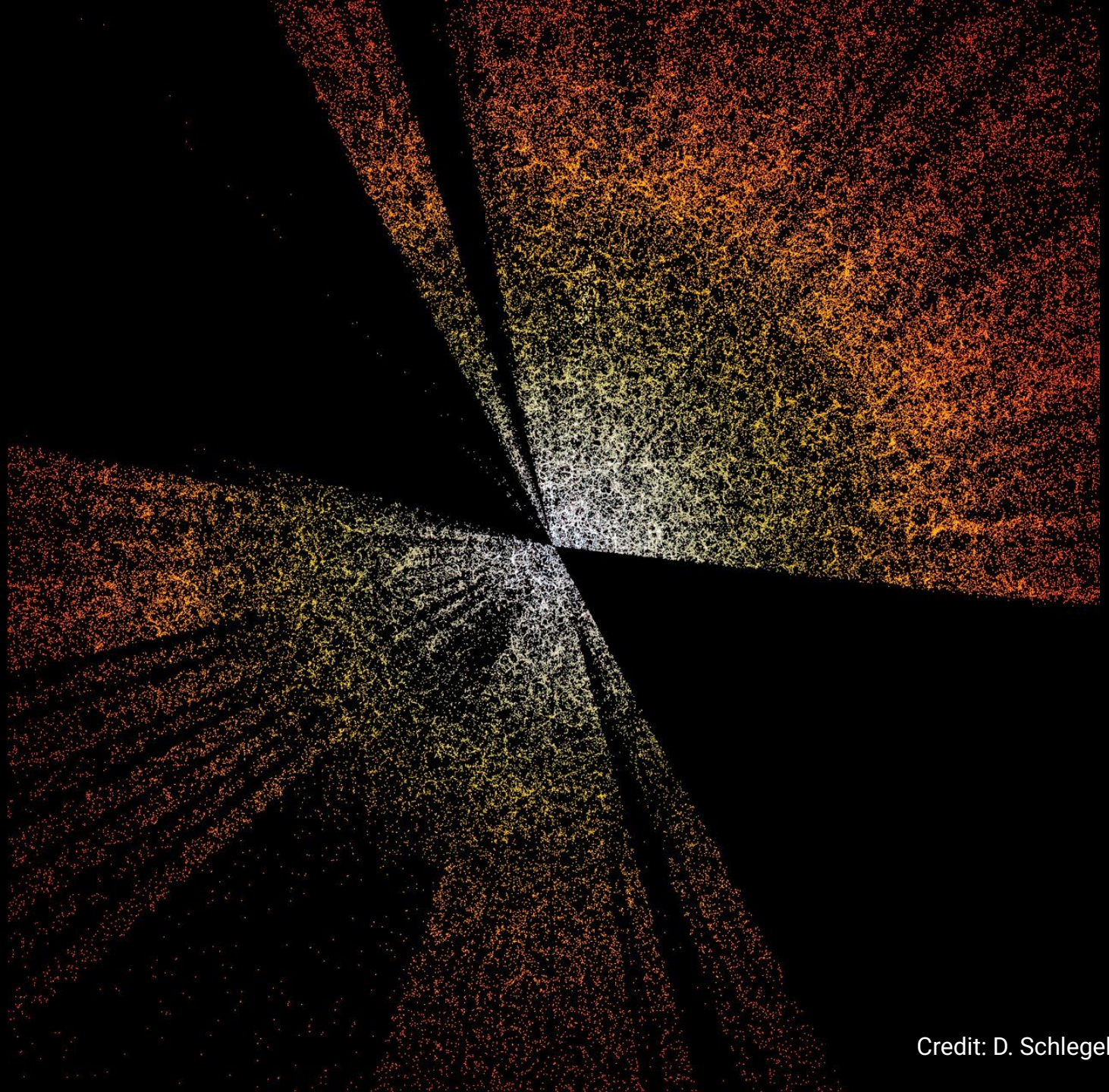
Backup Slides

From SDSS...



Credit: D. Schlegel/Berkeley Lab using data from DESI

...to DESI



Credit: D. Schlegel/Berkeley Lab using data from DESI

AURA