Data Lab
Open-Data, Open-Access Science Platform
Hands-On Tutorial

Robert Nikutta & Stephanie Juneau
(OIR Lab)

on behalf of the Data Lab team

AST 502 Science Platforms lecture & Data Lab tutorial

Tucson, March 16+18, 2020

Get your free Data Lab account: datalab.noao.edu

What is Data Lab today?

Soft side:
- ▶ Community of now ∼1200 users *(raison d'être)*
- ▶ Team of scientists, developers, DB architects, engineers, sysads *(make it work, enhance datasets, engage with users a lot!)*

Hard side:
- ▶ Repository for large (photometric) datasets *(catalogs & images)*
- ▶ Services to access data *(VO, SQL/ADQL, TAP, SIA, ...)*
- ▶ Gateway to other data centers
- ▶ Website / authenticated portal
- ▶ Exploratory tools *(survey coverage, catalog overlay, ...)*
- ▶ Cross-matching service
- ▶ Visualization *(all of Python, APIs)*
- ▶ Analysis facilities *(Jupyter notebook server, common-task helpers)*
- ▶ Remote storage *(vospace, mydb, notebook space)*

NSF's National Optical-Infrared
Astronomy Research Laboratory AURA

In the near future

- ▶ Science platform also for large spectroscopic data (e.g. DESI)
- ▶ Remote compute *(job manager)*
- ▶ Survey data publication facility *(BYOD)*
- ▶ Machine learning *(e.g. simple classifiers)*

Data Lab is a:
- ▶ Science Platform
- ▶ Data Exploration & Analysis environment
- ▶ Archive 2.0

(Pick your favorite *mot du jour*)

Why Data Lab?

- *Originally:* to exploit NOAO public data (4-m telescopes), surveys, time-domain; *specifically:* DES
- Bring user's analyses to the data
- Joint analysis with other datasets (ingested by us, or uploaded by user)
- Preparation of users for PB-scale data now, readiness for LSST
- *Today:* DES just one of 17 major surveys hosted at Data Lab

Some numbers on data holdings

NSF's OIR Lab Science Data Archive Holdings
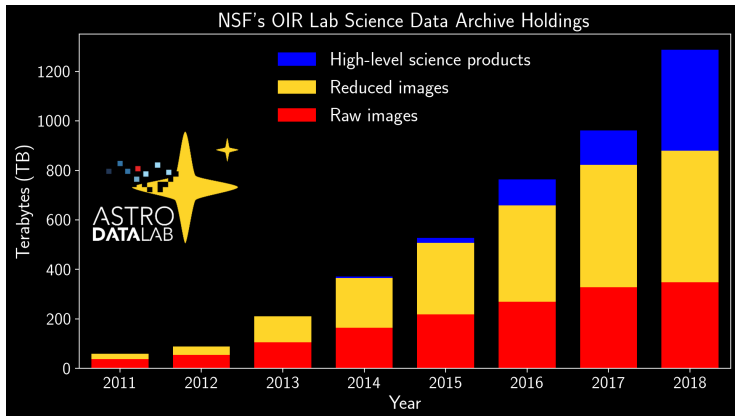
▶ As of now: ∼80 TB of catalog data, 150+ billion rows

▶ Images, masks, coadds, etc. (mostly at SDA): close to 5 PB

Some hosted datasets

Today (a selection):
- ▶ DES DR1 (400e6 objects)
- ▶ DECaLS DR 3, 4, 5, 6, 7, 8 (1.7e9 objects in DR8)
- ▶ DECaPS DR1 (2e9 objects)
- ▶ NOAO Source Catalog DR1 (2.9e9 objects, 30e9 measurements)
- ▶ Gaia DR1 and 2 (1.6e9 objects)
- ▶ PHAT (7.5e9 objects)
- ▶ AllWISE (750e6 objects)
- ▶ unWISE (2.2e9 objects)
- ▶ SMASH DR1 (100e6 objects) & DR2 (360e6 objects)
- ▶ SDSS (limited DR13 and 14)
- ▶ Stripe82 pre-cross-matched
- ▶ LSST simulated catalogs (7.2e9 objects)

Temporal information in: NSC, SMASH, LS DR 8-10, ...

Some hosted datasets

Future:

- ▶ DESI spectra
- ▶ DELVE
- ▶ skinny LSST
- ▶ skinny Pan-STARRS
- ▶ LS DR 9 & 10
- ▶ more of SDSS
- ▶ some X-ray catalogs
- ▶ Your data? (data publication service)

Next-gen wide-field cameras

Dark Energy Camera

Rubin/LSST Camera

570 Mpix

5.5'
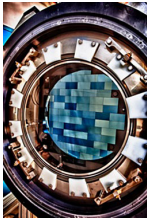
3200 Mpix

Sky coverage over time

2019.9
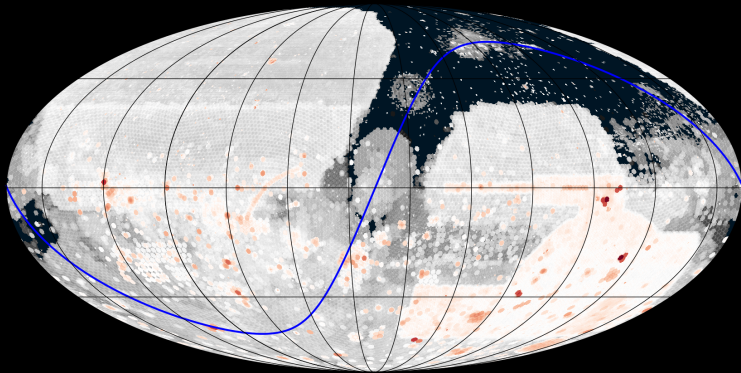
Exposure time at 2-4m telescopes in NSF's OIR Lab Science Data Archive

Current functionality

What	How
Authentication	web, `datalab` cmd line tool, `authClient.py`; *anonymous user exists*
Sky exploration	survey discovery tool (Aladin lite), catalog overlay (web)
Catalog query	web site, `queryClient.py`, `datalab`, TAP clients (e.g. TOPCAT)
Image query	Simple Image Access via website or notebooks (scriptable)
Query results...	to VOSpace, to mydb, or return to client
File transfer	`datalab`, `storeClient.py`
File services	collections of survey data, including suppl. files
Cross-matching	web tool, APIs
Visualization	notebooks, APIs
Analysis	Jupyter notebooks (all relevant packages present), or do locally
Common tasks	in `dl.helpers.*` (e.g. convert to pandas, sky plot, ...)

User = someone with a DL account

You...

- ▶ Get 1 TB of storage on vospace
- ▶ Get 250 GB of MyDB storage
- ▶ Can upload datasets for joint analysis
- ▶ Can edit/create/delete notebooks
- ▶ Can upload own Python source code
- ▶ Can share data with others
- ▶ Your workflows are persistent

Get your account today

Go to: `datalab.noao.edu` (click on 'Sign up')

ASTRO DATALAB · NSF · NSF's National Optical-Infrared Astronomy Research Laboratory · AURA

User community
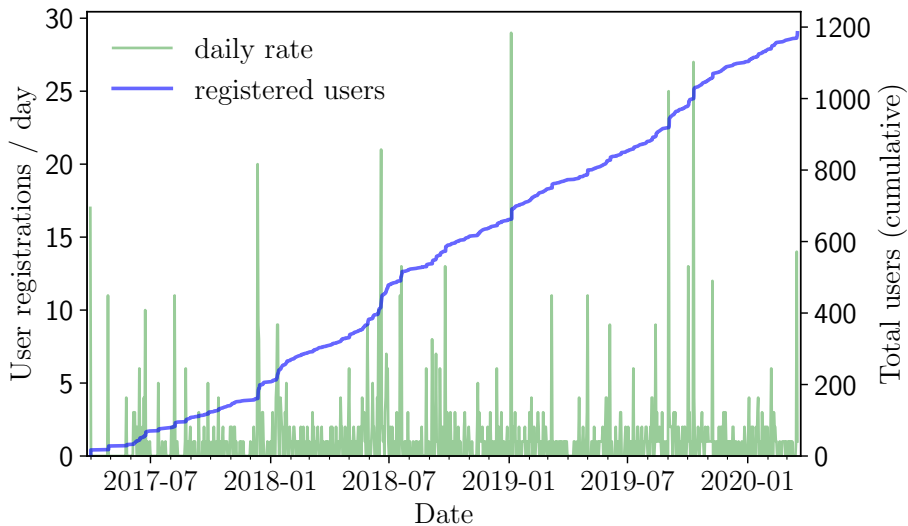
Curated default notebooks

Intros, How-Tos, Data Access, complete Science Cases, ANTARES filter dev kit, ...

Helpdesk & User Manual online

Ask a question (also anonymously),
FAQ, humans take care of you

Overviews, glossary, science cases,
SQL examples, tips & tricks

API documentation

Contact the Data Lab team

- datalab@noao.edu
- datalab.noao.edu
- github.com/noaodatalab
- @AstroDataLab

THANKS!
nikutta@noao.edu

Hacking together on great science with
Data Lab

Helpful links

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

Help desk
https://datalab.noao.edu/help/

How to sign-up and log-in
https://datalab.noao.edu/docs/manual/UsingTheNOAODataLab/WebInterfaces/
CreatingAnAccountLoggingIn/CreatingAnAccountLoggingIn.html

On the Jupyter notebook server, and Data Lab example/how-to notebooks
https://datalab.noao.edu/docs/manual/UsingTheNOAODataLab/JupyterNotebooks/
JupyterNotebooks.html

How to copy latest Data Lab notebooks to a new (writable) directory
https://datalab.noao.edu/docs/manual/UsingTheNOAODataLab/JupyterNotebooks/
JupyterNotebooks.html#get-the-latest-set-of-default-notebooks

Notebooks on Github (check out README and CONTRIBUTING files)
https://github.com/noaodatalab/notebooks-latest/