



# The NOAO Data Lab

**Robert Nikutta**

on behalf of the NOAO Data Lab team

*Data Visualization and Exploration in the LSST Era*  
*NCSA, June 2018*

# What Data Lab is...

## Soft side:

- ▶ Community of 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 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, ...*)
- ▶ Visualization (*all of Python, APIs*)
- ▶ Analysis facilities (*Jupyter notebook server, common-task helpers*)
- ▶ Remote storage (*vospace, mydb, notebook space*)

# What Data Lab is...

Near future:

- ▶ Remote compute (*job manager*)
- ▶ Cross-matching service
- ▶ Machine learning (*e.g. simple classifiers*)
- ▶ Survey data publication facility (*BYOD*)

Data Lab is a:

- ▶ Science Platform
- ▶ Data Exploration & Analysis environment
- ▶ Archive 2.0

(Pick your favorite *mot du jour*)

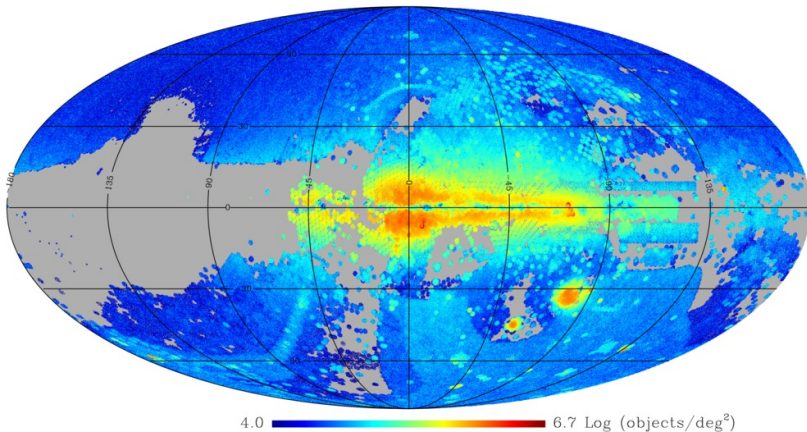


# Why Data Lab?

- ▶ Exploit NOAO public data (4-m telescopes), surveys, time-domain
- ▶ 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
- ▶ As of now:  $\sim 45$  TB of catalog data, 100+ billion rows
- ▶ With images, masks, etc.: 0.5+ PB

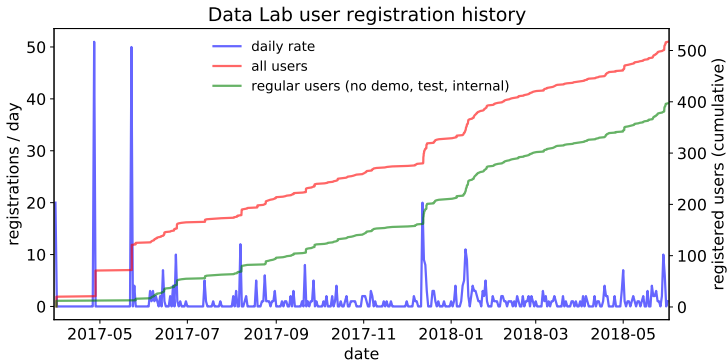


# Log number of objects / $\text{deg}^2$ in the NSC



PI: D. Nidever

# User community



- ▶ Helpdesk, FAQ, humans take care of you
- ▶ Data Lab user manual online  
(overviews, glossary, science cases, SQL examples, tips & tricks)
- ▶ API documentation

# Hosted datasets

Today (a selection):

- ▶ LS DR 3, 4, 5, 6 (1e9 objects in DR5+6)
- ▶ DES DR1 (400e6 objects)
- ▶ NOAO Source Catalog (2.9e9 objects, 30e9 measurements)
- ▶ Gaia DR1 and 2 (1.6e9 objects)
- ▶ PHAT
- ▶ DECaPS DR1 (2e9 objects)
- ▶ ALLWISE
- ▶ SMASH (100e6 objects)
- ▶ SDSS (limited DR13 and 14)
- ▶ Stripe82 pre-cross-matched

Temporal information:

- ▶ NSC, LS DR7(?), SMASH, ...



# Hosted datasets

## Future:

- ▶ DESI spectra
- ▶ LSST simulated catalogs
- ▶ skinny Pan-STARRS
- ▶ skinny LSST
- ▶ LS DR 7
- ▶ more of SDSS
- ▶ some X-ray catalogs
- ▶ Your data? (data publication service)



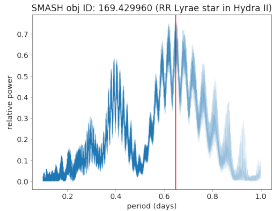


# Current functionality

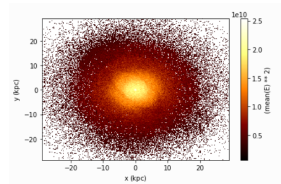
What	How
Authentication	web, datalab cmd line tool, <code>authClient.py</code> ; <i>anonymous user exists</i>
Sky exploration	survey discovery tool (Aladin lite), catalog overlay (web)
Catalog query	web site, <code>queryClient.py</code> , datalab, TAP clients (e.g. TOPCAT)
Image query	Simple Image Access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

# Visualization examples

matplotlib



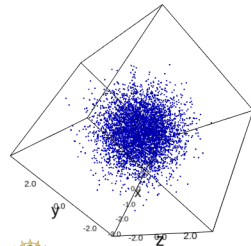
vaex



ipyvolume / ipywidgets

Also

- ▶ bokeh
- ▶ exploring Glue
- ▶ Firefly
- ▶ ...



# Contact us

- ▶ [datalab@noao.edu](mailto:datalab@noao.edu)
- ▶ [datalab.noao.edu](http://datalab.noao.edu)
- ▶ [github.com/noao-datalab](https://github.com/noao-datalab)
- ▶ [@NOAODataLab](https://twitter.com/NOAODataLab)

THANKS!

[nikutta@noao.edu](mailto:nikutta@noao.edu)

**Hands-on today 2:15–3:30pm  
in room 1040!**



Data Lab call center agents ready for your call