

# The NOAO Data Lab

**Robert Nikutta**

on behalf of the NOAO Data Lab team

*Near-Field Cosmology with the  
Dark Energy Survey's DR1 and Beyond*

*University of Chicago, 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



# What does a user account get you?

[Register as a new user](#)

Username:

Must use between 6 and 30 characters, only lower case letters, numbers and underscores allowed.

Password:

Must use 6 or more characters, and must contain UPPER/lowercase and numbers.

Full Name:

Email:

Privacy: Your email address will not be shared or sold to third parties.

Affiliation:

Anti-spam verification:



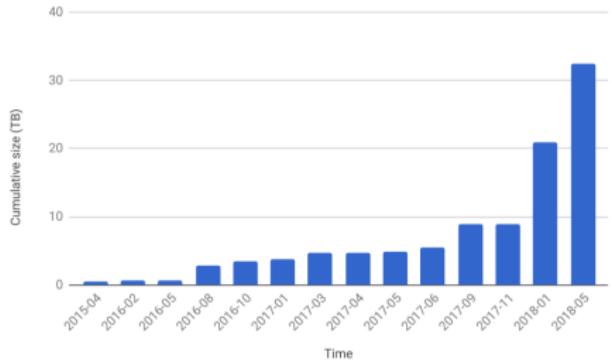
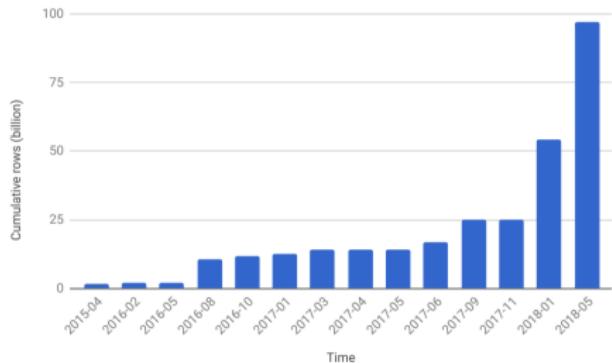
[Register](#)

- ▶ You get 1 TB of storage on vospace (soft quota)
- ▶ You get 100 GB of MyDB storage (soft quota)
- ▶ You can upload even large datasets for joint analysis
- ▶ You can edit/create/delete notebooks
- ▶ You can upload own Python source code
- ▶ Your workflows are persistent

## Get your account today

Go to: [datalab.noao.edu](http://datalab.noao.edu) (click on 'Sign up')

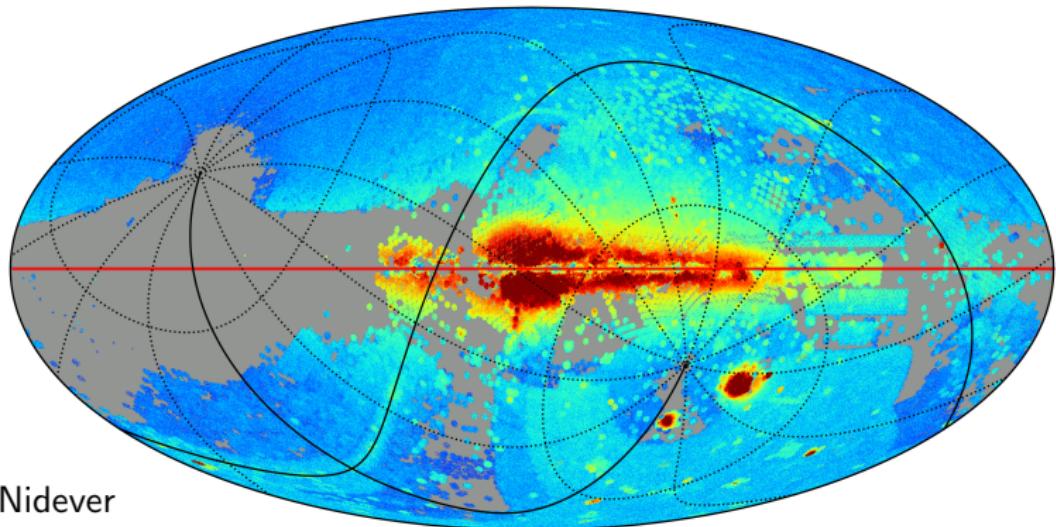
# Some current numbers



- ▶ As of now: ~45 TB of catalog data, 100+ billion rows
- ▶ With images, masks, etc.: 0.5+ PB

# NOAO Source Catalog

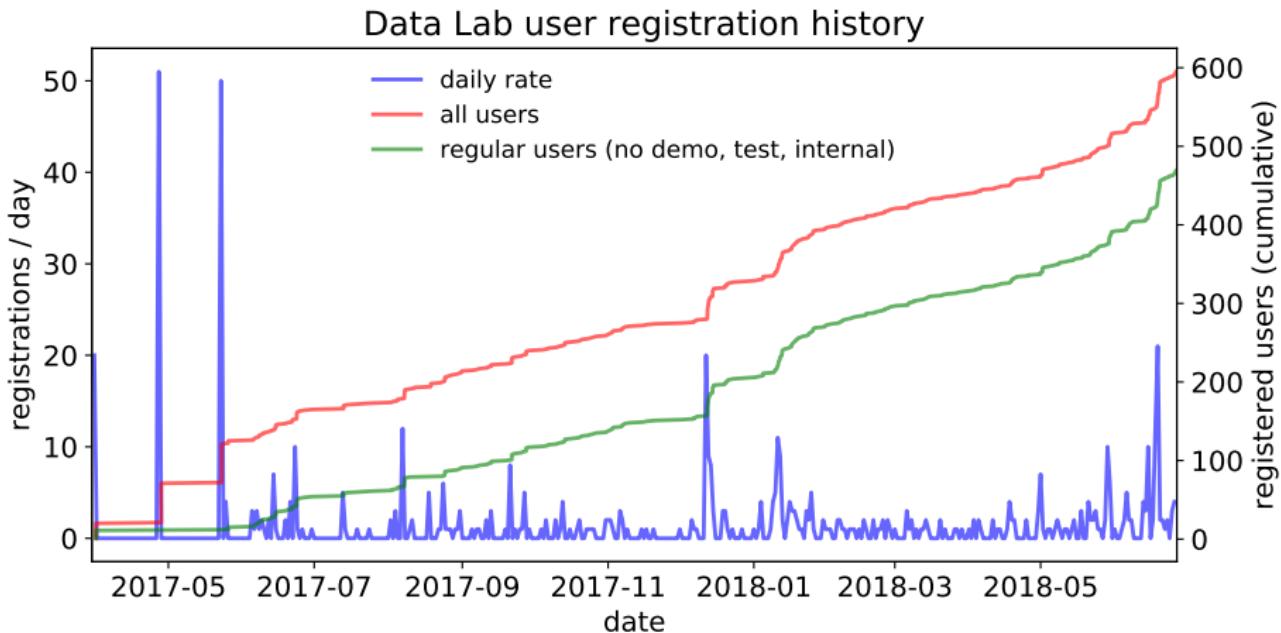
Log number of objects / deg<sup>2</sup> in the NSC



PI: D. Nidever

- ▶ 2.9e+9 sources
- ▶ 30e+9 measurements

# User community







# Helpdesk

The screenshot shows the Data Lab User Forum interface. At the top, there's a navigation bar with File, Edit, View, Tools, Window, Help, and a Data Lab Help Desk section. Below the navigation is a header with the NOAO Data Lab User Forum logo and links for Forum home, All Activity, Questions, Unanswered, Tags, Categories, Users, and Ask a Question.

The main area features a large globe map with blue and green data points. Overlaid on the globe is a banner with the text "Data Lab User Forum". Below the globe are three boxes: "Search answers for all your queries" (with a magnifying glass icon), "One destination for all your queries" (with a question mark icon), and "Get answers from the experts" (with a checkmark icon).

Below the globe map, there's a section titled "Recent questions and answers" with a "Recent" button. It includes a search bar labeled "Ask a question:" and a list of recent posts. One post is highlighted: "Unable to login to Datalab" by karen (300 points) | 8 views | answered 5 days ago | datab...". To the right of the post are "0 votes" and "1 answer". A search bar with a magnifying glass icon is also present.

On the right side of the interface, there are statistics: "68 questions" and "64 answers".

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

# Data Lab user manual online

The screenshot shows a web browser window displaying the "NOAO Data Lab documentation". The title bar reads "File Edit View Tools Window Help Data Lab documentation — NOAO Data Lab next | index". The main content area has a header "Welcome to the NOAO Data Lab documentation". Below it is a "Contents" section with a hierarchical menu:

- 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

Below the contents is a "Indices and tables" section with a list:

- Index
- Module Index
- Search Page

The footer contains the text "NOAO Data Lab 1.1.0 documentation » next | index" and "© Copyright 2016-2017, NOAO Data Lab <data@noao.edu>. Last updated on 2018-Jan-07."

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



# API documentation

File Edit View Tools Window Help documentation (auto-generated) — datalab · next | modules | index

datalab 1.1.0 documentation »

## Table Of Contents

- NOAO Data Lab API documentation (auto-generated)
- Indices and tables

### Next topic

authClient module

### This Page

Show Source

### Quick search

Go

## NOAO Data Lab API documentation (auto-generated)

### Contents:

- authClient module
- queryClient module
- storeClient module
- helpers module
  - cluster helpers
  - crossmatch helpers
  - plot helpers
  - utils helpers
  - legacy helpers
  - all helpers
- dlinterface module
- dtasks module

## Indices and tables

- Index
- Module Index
- Search Page

next | modules | index

© Copyright 2017, DataLab team. Last updated on 2017-Dec-20. Created using Sphinx 1.6.3.

File Edit View Tools Window Help authClient module — datalab 1.1.0 documentation · previous | next | modules | index

datalab 1.1.0 documentation »

### Previous topic

queryClient module

### This Page

Show Source

### Quick search

Go

## authClient module

`class authClient.authClient` [source]

**Bases:** `object`

**AUTHCLIENT** – Client-side methods to access the Data Lab Authentication Service.

**debug(debug\_val)** [source]

**get\_profile()** [source]

Get the requested service profile.

**Parameters:** `None` –

**Returns:** `profile` – The currently requested service profile.

**Return type:** `str`

**Example**

```
from dl import authMgr
profile = authMgr.client.get_profile()
```

**get\_service()** [source]

Return the currently-used Authentication Service URL.

**Parameters:** `None` –

**Returns:** `service_url` – The currently-used Authentication Service URL.

**Return type:** `str`

**Example**

```
from dl import authMgr
service_url = authMgr.client.get_service()
```

**hasAccess(user, resource)** [source]



# Hosted datasets

Today (a selection):

- ▶ **DES DR1 (400e6 objects)** [data also at NCSA & LINeA (Brazil)]
- ▶ DECaLS DR 3, 4, 5, 6 (1e9 objects in DR5+6)
- ▶ DECaPS DR1 (2e9 objects)
- ▶ NOAO Source Catalog (2.9e9 objects, 30e9 measurements)
- ▶ Gaia DR1 and 2 (1.6e9 objects)
- ▶ PHAT
- ▶ 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)

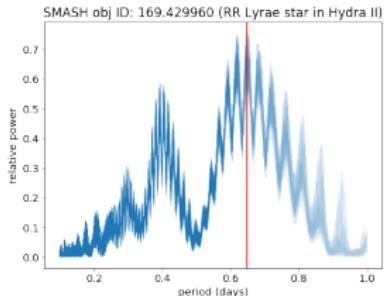


# Current functionality

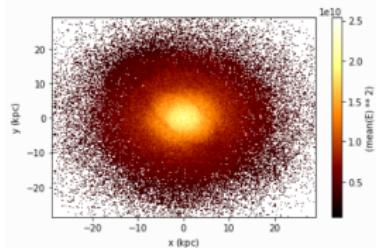
What	How
Authentication	web, datalab cmd line tool, authClient.py; <i>anonymous user exists</i>
Sky exploration	survey discovery tool (Aladin lite), catalog overlay (web)
Catalog query	web site, queryClient.py, datalab, TAP clients (e.g. TOPCAT)
Image query	Simple Image Access via website or notebooks (scriptable)
Query results...	to VOSpace, to mydb, or return to client
File transfer	datalab, storeClient.py
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, ...)

# Visualization examples

matplotlib



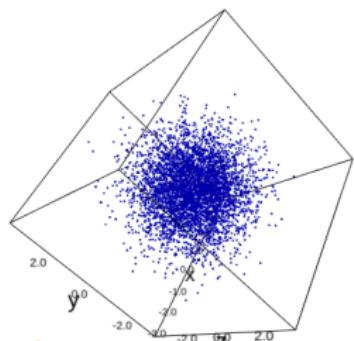
vaex



ipyvolume / ipywidgets

Also

- ▶ bokeh
- ▶ exploring Glue, Firefly
- ▶ spectra visualization tools  
(e.g. specviz)
- ▶ ...



# 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)

Data Lab Hands-On

This Friday 9–10am

Get an account today!

(and hack-together?)



Data Lab call center agents  
ready for your call