



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*)

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:

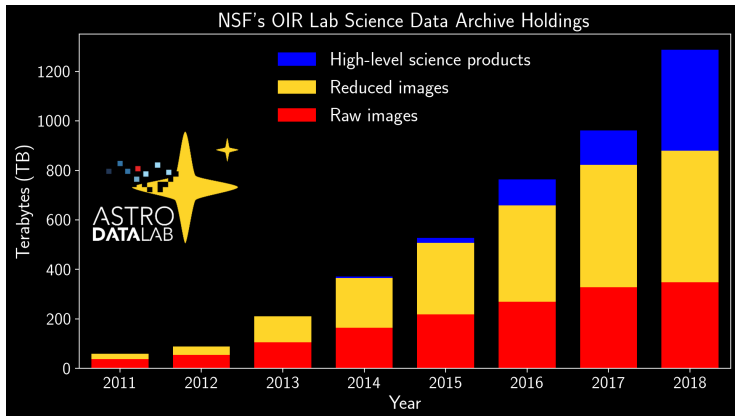
- ▶ Science Platform
- ▶ Data Exploration & Analysis environment
- ▶ Archive 2.0

(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



- ▶ 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, ...

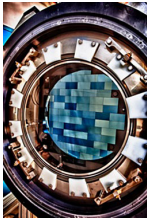
Some hosted datasets

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



570 Mpix

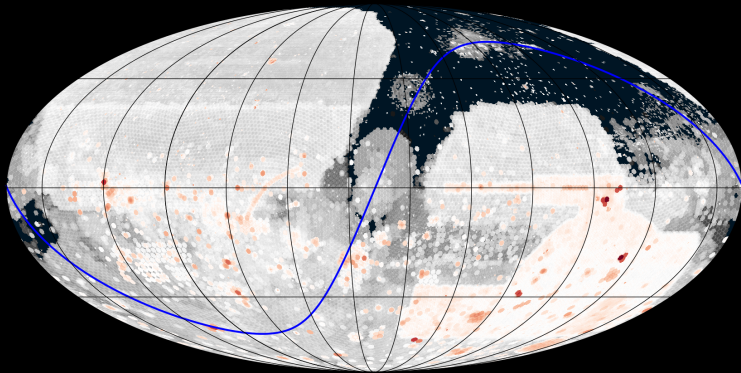
Rubin/LSST Camera



3200 Mpix

Sky coverage over time

2019.9



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



NSF's National Optical-Infrared
Astronomy Research Laboratory



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	<code>datalab</code> , <code>storeClient.py</code>
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 <code>dl.helpers.*</code> (e.g. convert to pandas, sky plot, ...)

User = someone with a DL account

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:

☐ I'm not a robot


reCAPTCHA
[Privacy](#) [Terms](#)

 Register

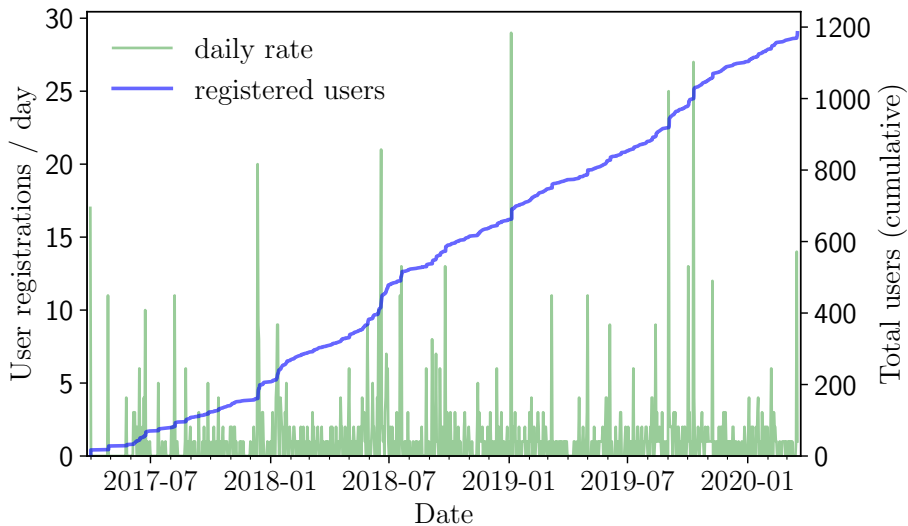
You...

- ▶ 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')

User community



Curated default notebooks

noaodatalab / notebooks-latest

Unwatch 7 Star 1 Fork 6

Code Issues 3 Pull requests 0 Projects 0 Wiki Security Insights Settings

Default set of Data Lab notebooks, by DL team and contributed by users [Edit](#)

[Manage topics](#)

149 commits 1 branch 0 releases 11 contributors BSD-3-Clause

Branch: master New pull request Create new file Upload files Find file Clone or download

rmikutta Merge pull request #18 from gnarayan/master Latest commit 42d6b04 2 days ago

01_GettingStartedWithDataLab	Minor updates to 02_GettingStartedWithDataLab.ipynb and DwarfGalaxies...	6 months ago
02_DataAccessOverview	Consolidation: restructure, remove leading _, remove _yyymmdd suffixes.	9 months ago
03_ScienceExamples	Minor updates to 02_GettingStartedWithDataLab.ipynb and DwarfGalaxies...	6 months ago
04_HowTos	removed empty notebook cell per PR comment	3 months ago
05_Contrib	Latest version of ANTARES DevKit with updated examples	8 days ago
06_EPO	Revert change to Teen Astronomy Cafe	2 days ago
.gitignore	update gitignore	10 months ago
CONTRIBUTING	- Move 05_Contrib/CONTRIBUTE to the more standard location ./CONTRIBU...	9 months ago
DataLabNotebookTemplate.ipynb	- Add 05_Contrib/CONTRIBUTE file on how to contribute notebooks to DL	9 months ago
LICENSE	Create LICENSE	9 months ago
README.txt	Update README.txt with info about 06_EPO/	3 months ago
keywords.txt	Sort keywords.txt	9 months ago

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

Helpdesk & User Manual online

The screenshot shows the 'Data Lab Help Desk' window. The main content area features a large, colorful astronomical image of a galaxy or nebula. Overlaid on this image is the 'Data Lab User Forum' title and three search buttons: 'Search answers for all your queries', 'One destination for all your queries', and 'Get answers from the experts'. Below the image, there is a section for 'Recent questions and answers' with a search bar and a list of questions. One question is visible: 'Unable to login to DataLab' with 0 votes and 1 answer. The interface includes a standard web browser menu at the top and a sidebar with navigation links like 'Forum home', 'All Activity', 'Questions', 'Unanswered', 'Tags', 'Categories', 'Users', and 'Ask a Question'.

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

The screenshot shows the 'NOAO Data Lab 1.1.0 documentation' window. The main content area is titled 'Welcome to the NOAO Data Lab documentation' and features a large, colorful astronomical image of a galaxy or nebula. Below the image, there is a 'Table Of Contents' section with links to 'Welcome to the NOAO Data Lab documentation', 'Indices and tables', 'Next topic', 'This Page', 'Show Source', and 'Quick search'. The 'Table Of Contents' section lists various topics, including '1. Using the NOAO Data Lab', '2. FAQs', '3. Appendices', and '4. Data Publication docs'. The 'Indices and tables' section lists 'Index', 'Module Index', and 'Search Page'. The interface includes a standard web browser menu at the top and a sidebar with navigation links like 'next' and 'index'.

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



NSF's National Optical-Infrared
Astronomy Research Laboratory



API documentation

datalab 1.1.0 documentation » next | modules | index

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
- dinterface module
- dltasks module

Indices and tables

- Index
- Module Index
- Search Page

datalab 1.1.0 documentation » next | modules | index

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

datalab 1.1.0 documentation » previous | next | modules | index

authClient module

class authClient.**authClient** [\[source\]](#)
Bases: **object**

Next topic
queryClient module

This Page
Show Source

Quick search

Go

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

debug(*debug_val*) [\[source\]](#)
Get the requested service profile.

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*) [\[source\]](#)



NSF's National Optical-Infrared
Astronomy Research Laboratory



Contact the Data Lab team

- ▶ datalab@noao.edu
- ▶ datalab.noao.edu
- ▶ github.com/noaodatalab
- ▶ [@AstroDataLab](https://twitter.com/AstroDataLab)

THANKS!

nikutta@noao.edu



Hacking together on great science with
Data Lab



NSF's National Optical-Infrared
Astronomy Research Laboratory



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/>