



# *The NOAO Data Lab*

## *Design, Capabilities and*

## *Community Development*

Michael Fitzpatrick  
for the Data Lab Team



**National Optical Astronomy Observatory**

Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# What is it?

Data Lab is **Science Exploration Platform** that provides:

- ◆ Repository for large datasets (*catalogs & their parent images*)
- ◆ Data access services (*VO, SQL/ADQL, TAP, SIA,...*)
- ◆ Virtual Storage (*VOSpace, MyDB, notebooks*)
- ◆ Gateway to remote data centers
- ◆ Data Sharing / Collaboration / Publication
- ◆ Website / authenticated portal / documentation
- ◆ Exploratory tools (*survey coverage, catalogs overlays, ...*)
- ◆ Visualization tools (*all-sky browser, custom APIs*)
- ◆ Analysis facilities (*Jupyter, utility methods, Compute Service*)



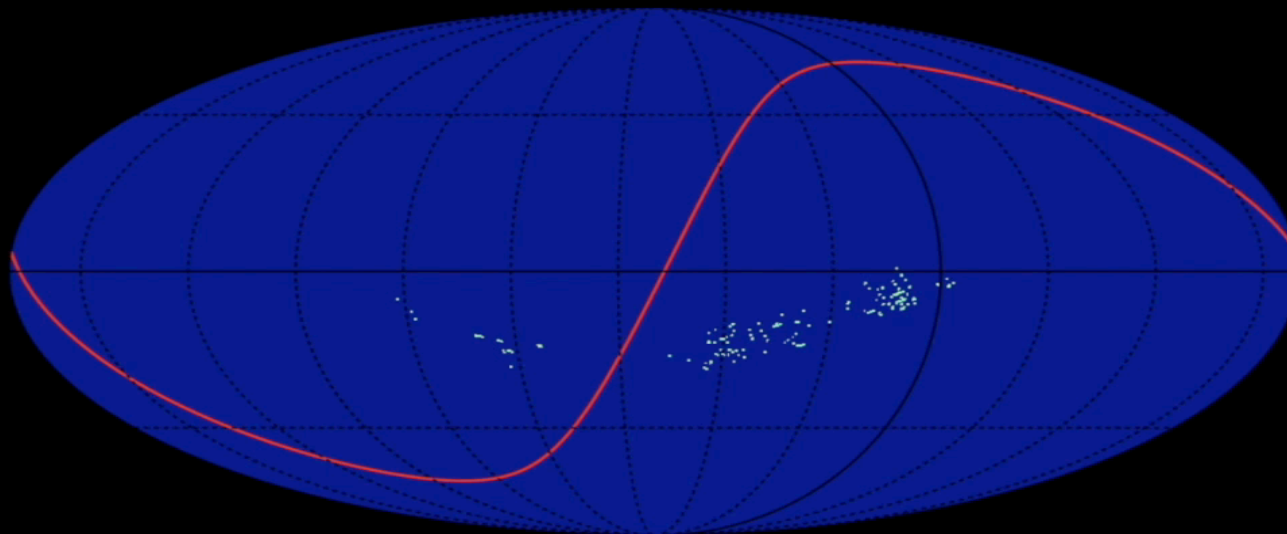
National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# Project Motivation

2004

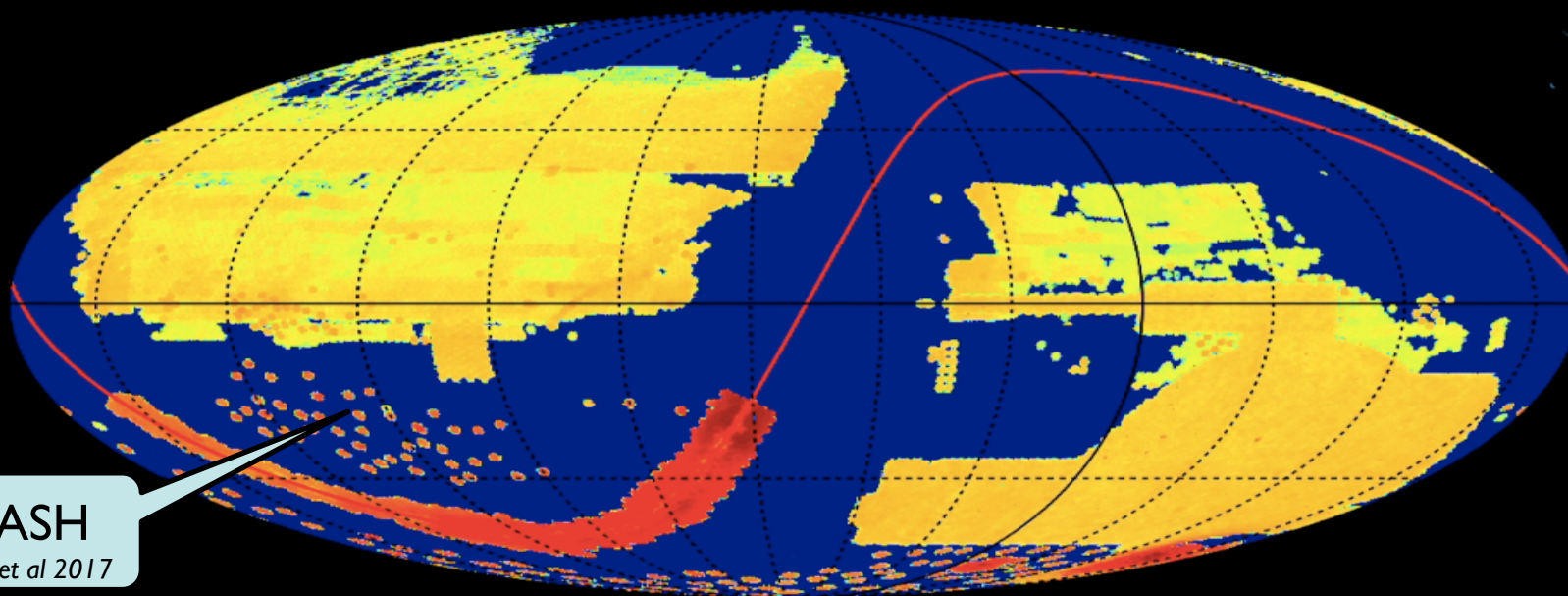


**National Optical Astronomy Observatory**  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center



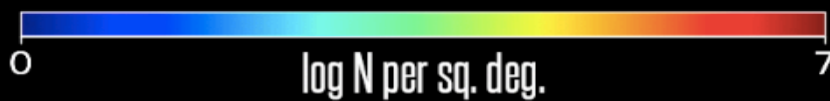


# SMASH, DECaPS, Legacy Survey, DES



**SMASH**

*Nidever et al 2017*

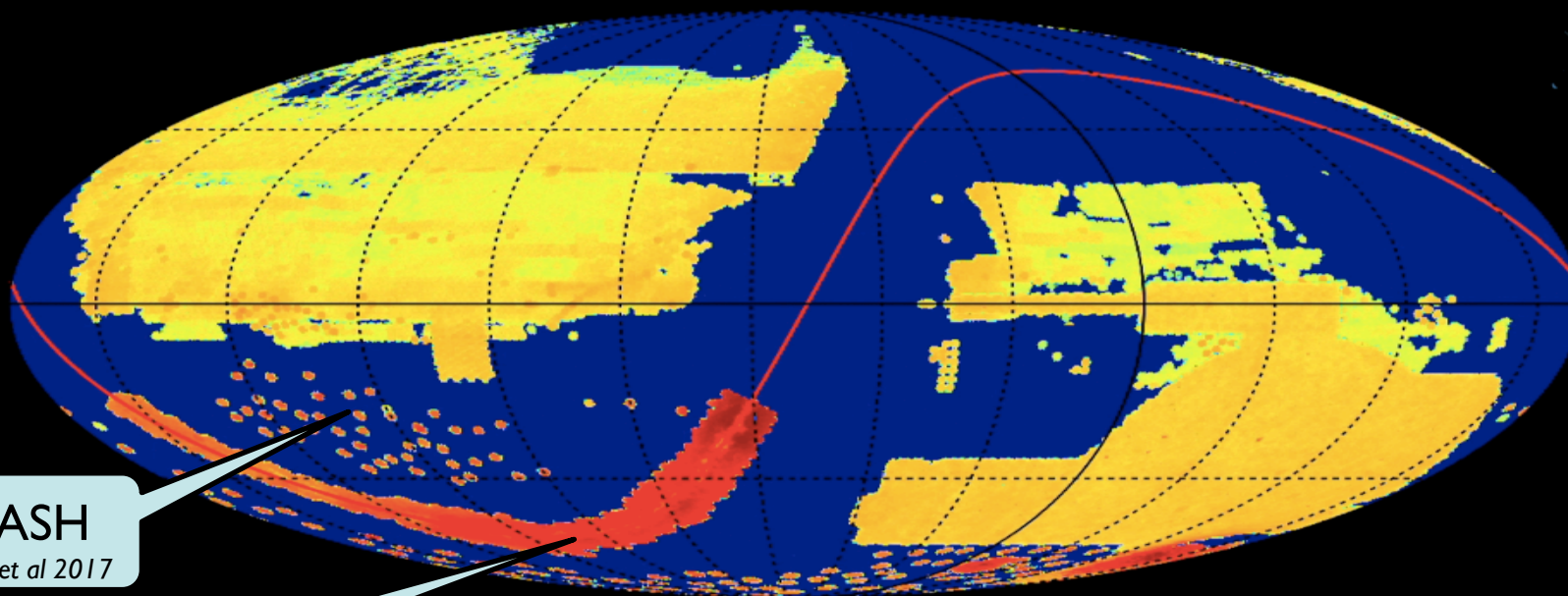


National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# SMASH, DECaPS, Legacy Survey, DES

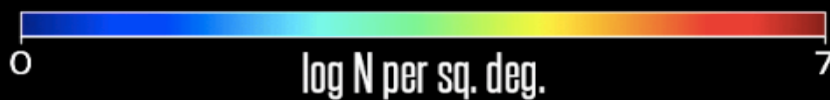


**SMASH**

*Nidever et al 2017*

**DECaPS**

*Schlafly et al 2017*



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# SMASH, DECaPS, Legacy Survey, DES

**DECaLS**

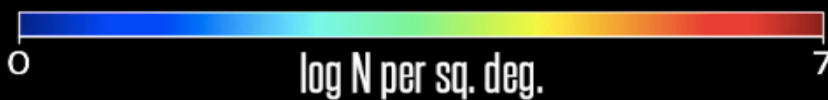
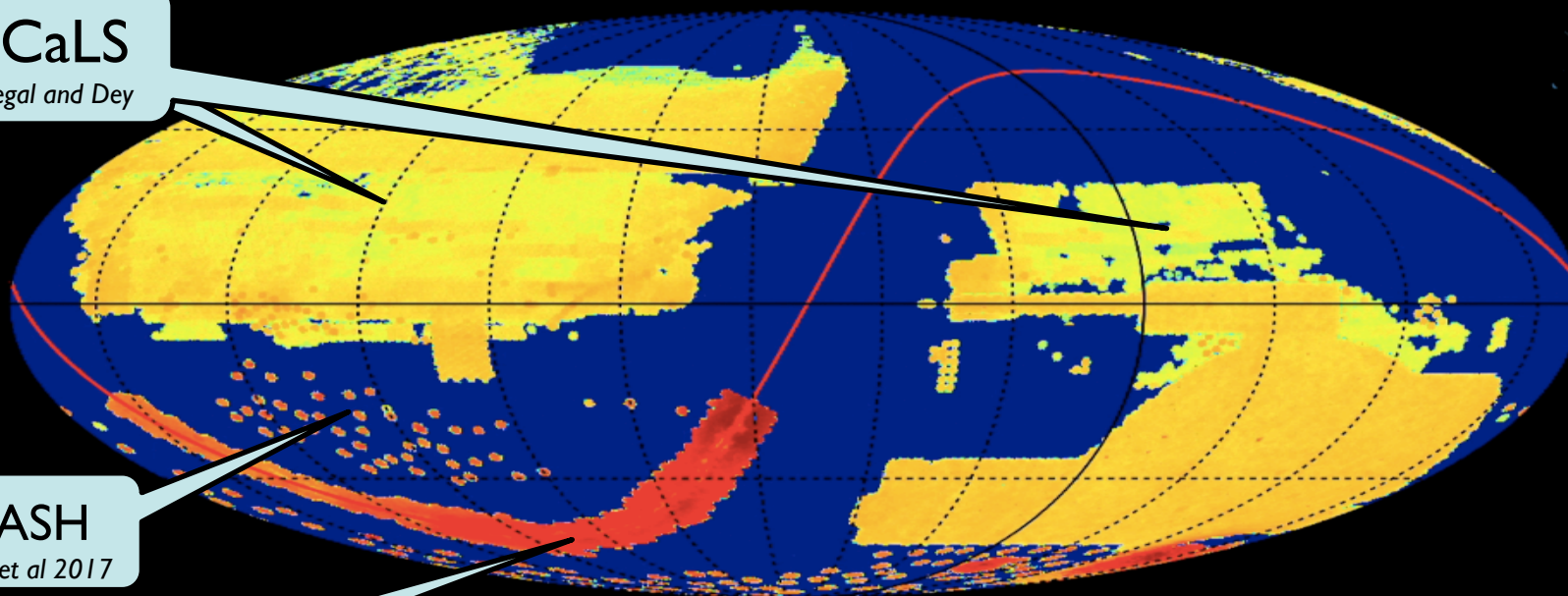
*PI: Schlegel and Dey*

**SMASH**

*Nidever et al 2017*

**DECaPS**

*Schlafly et al 2017*



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# SMASH, DECaPS, Legacy Survey, DES

**DECaLS**

PI: Schlegel and Dey

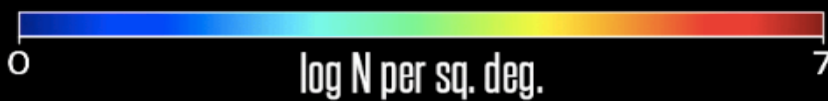
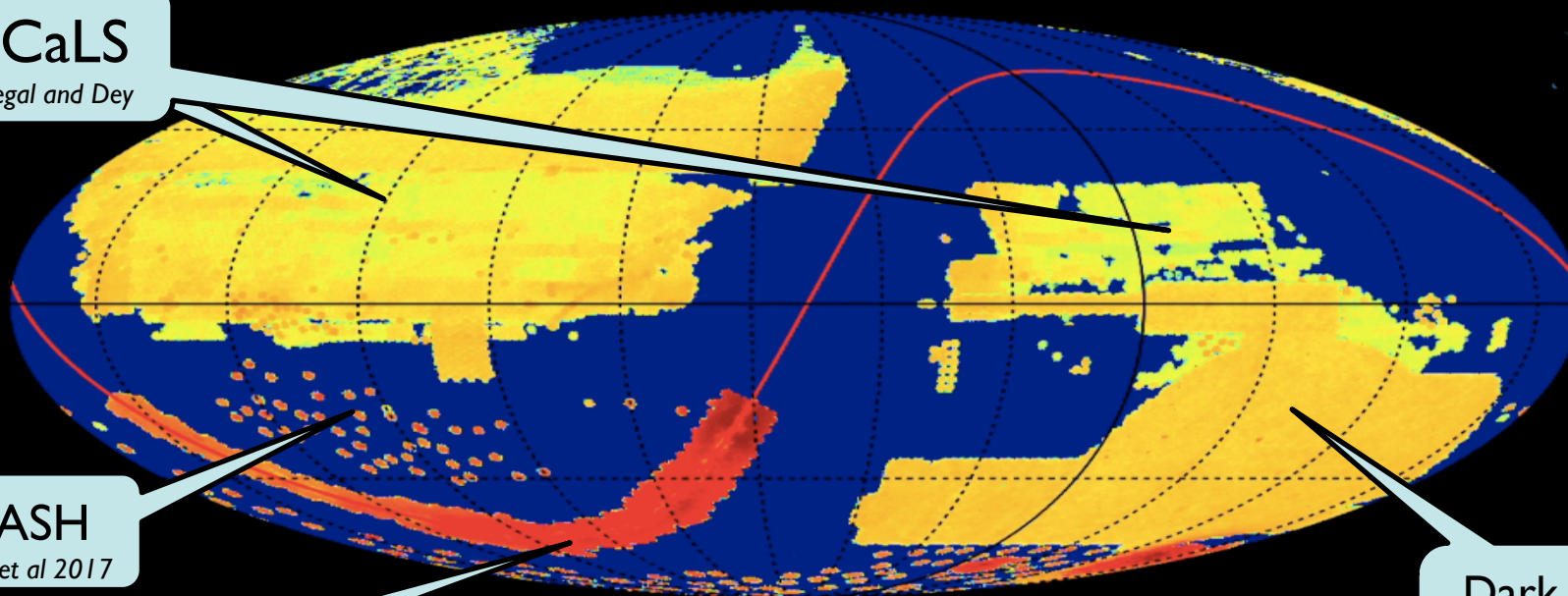
**SMASH**

Nidever et al 2017

**DECaPS**

Schlafly et al 2017

**Dark  
Energy  
Survey**

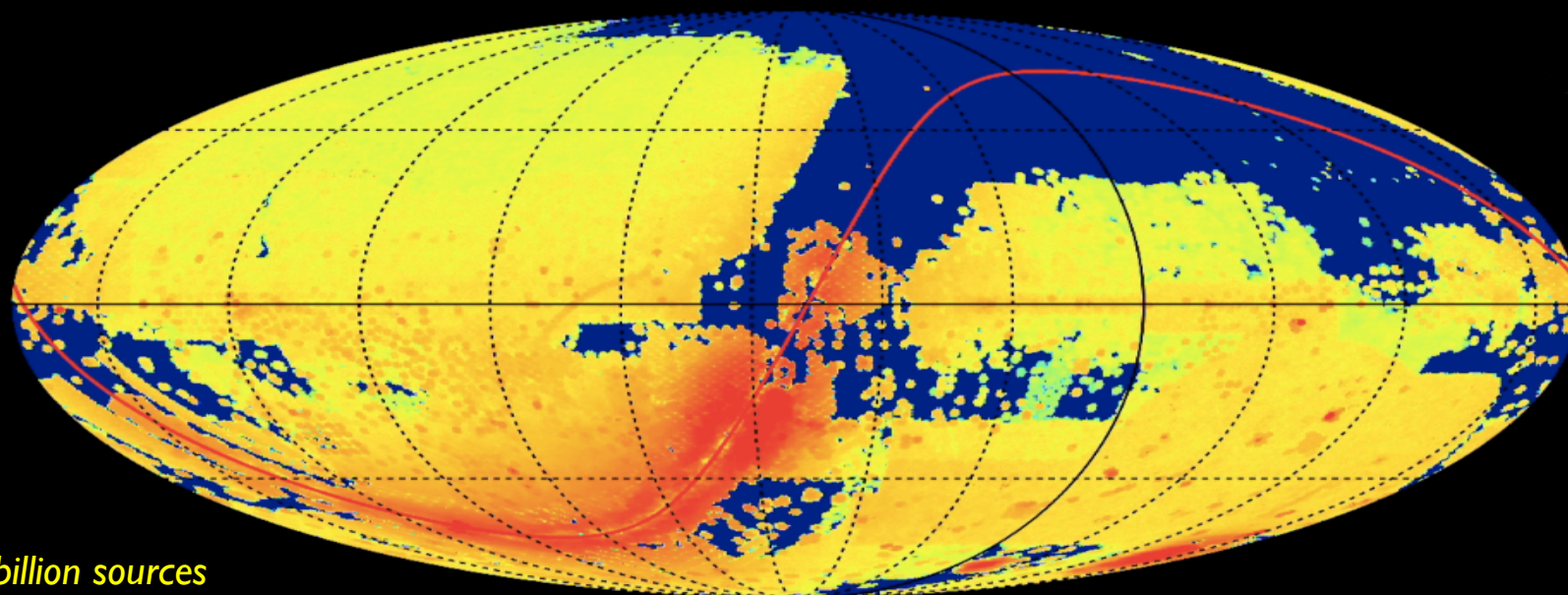


National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center

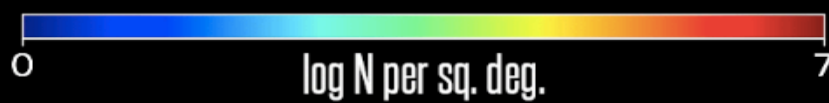




# NOAO Source Catalog



2.9 billion sources  
30 billion measurements



*Nidever et al (2018)*



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center







# Data Holdings

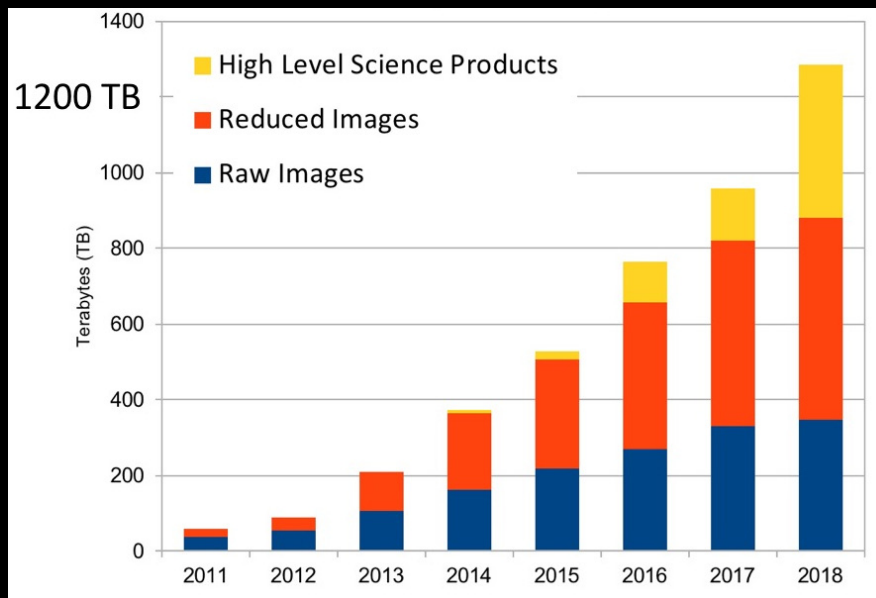
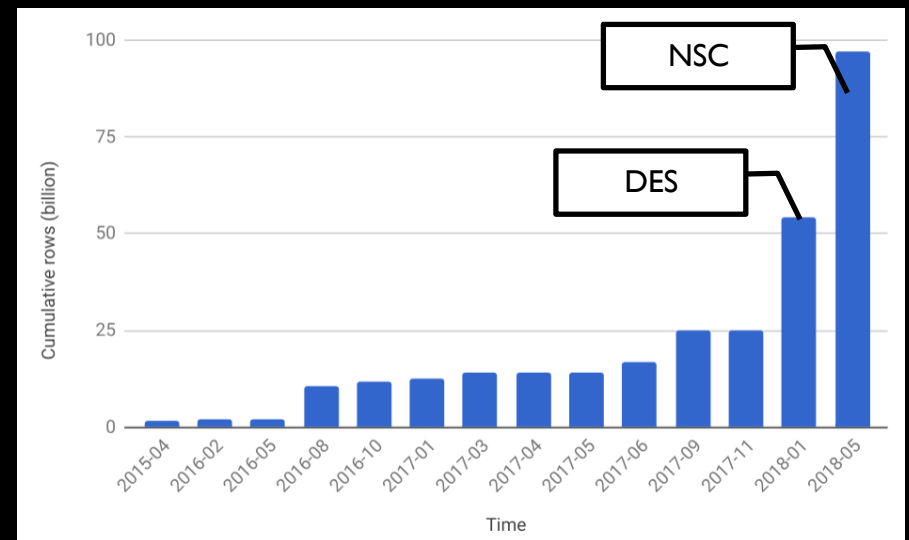


Image Archive Holdings (~1.2PB)

## Catalog Database (100B rows, ~50TB)



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# Data Holdings

## Catalog Data Holdings

- 316 tables in 38 different schema
  - 15 different image services (by survey, reduction type)
  - **Major Surveys:** DES, DESI, DECaLS, DECaPS, PHAT, S-PLUS, NSC, ....
  - **Ref data:** AllWISE, Gaia, USNO, 2MASS, SDSS, .....
  - **Planned(?) Datasets:** CRTS, Vista HS, UKIDSS, Chandra/XMM XSC, ....
- 
- Q3C spatial indexing + data clustering + type ordering for performance
  - TAP access to all tables (including image metadata)
  - Cone Search services for all tables w/ positional data
  - Image cutout services (SIA)



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center

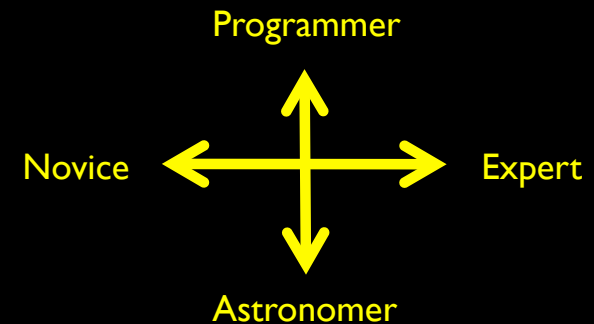




# Design Principles to Consider

## Know your audience

- Astronomers and Programmers will approach your system in different ways
- They will have a range of experience and skills
- A good UI will be inclusive



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# Design Principles to Consider

## Allow multiple entry points into the system

- Web-based Portals
- Notebook Servers
- Command-line Tools
- Programmatic APIs

Novice User / Casual Browsing



Blackbelt User / Workflow Dev



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





## Design Principles to Consider

Enable *user-developed* tools, don't just supply them

- Support legacy code where possible
- Publish/Share code and notebooks
- Host a repo to support user contributions
- Allow for user scripting and workflow development



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# Design Principles to Consider

## Use established standards, but hide complexity

- Provide *Astronomer Friendly* interfaces
- Provide *Low-Level* interfaces
- Realize these are *separate* things (and that's okay) !
- Most of all: **Enable Science!**



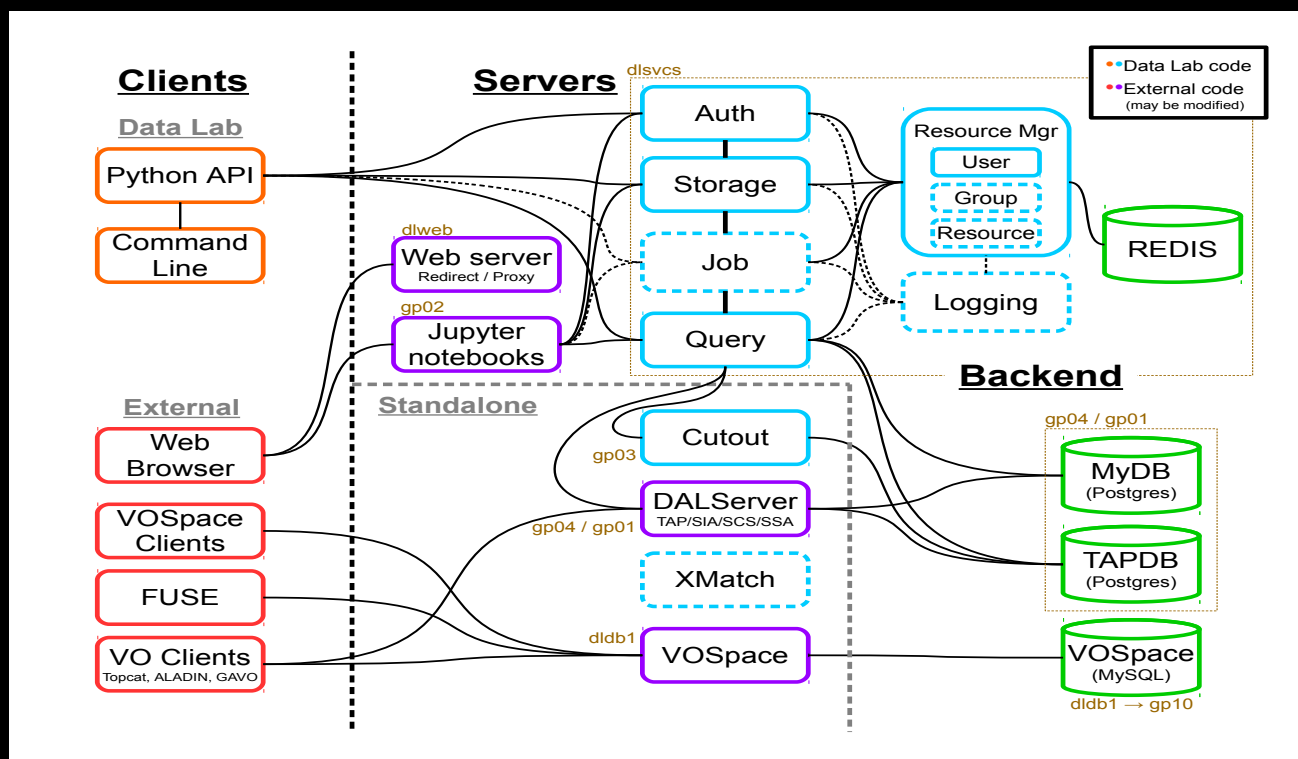
National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center

AURA





# Architecture/Service Overview



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# VO Inside

## VO Protocols and Data Access

- **VOSpace** for Virtual Storage
- **TAP** for Catalog Access
  - **ADQL** query language (custom funcs in development)
- **SCS** for Catalog Access
- **SIA** for Image Access
  - SIA v1 in use, SIA v2 available
  - **ObsCore** data model for images
- **SSA** available for Spectral Access (not currently used)
- **VOTable** support in client interfaces
- **HiPS** for Discovery Tool base image layers (MOC use planned)
- **Universal Worker Service** (planned for Compute Service)
- **VOSI** for VOSpace/TAP/Manager services
- **OAI Publishing Registry** for services planned



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center







# Public Data Access

- Access to public image / catalog data does not require an account
- Anonymous Jupyter notebook server allows a 'Trial Run' of Data Lab before creating an account
  - Scratch/demo notebooks, no persistence beyond session
  - Containerized environment initialized for each new user
- Downloadable *datalab* client and API allow anonymous desktop and programmatic access to Data Lab tools and services
- Web tools (*Data Discovery*, *Query Interface*, etc) can use anonymous access



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# User Accounts

## User Account Creation

- Sign-up form for account request on web site
- New accounts require approval before resources are allocated
- Currently open to professional researchers only
- *Future plans for*
  - *EPO/student accounts (e.g. custom notebooks)*
  - *Group accounts*
  - *Limited public access using temporary storage/ compute services*

### 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](#)



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# User Accounts

## Resource Allocations

- Virtual disk storage: 1 TB
- MyDB database storage: 250 GB
- Jupyter notebook storage: *no hard limit, but not unlimited*
  
- Sync query timeout: 120 sec (*max: 600 sec*)
- Async query timeout: 24-hrs (*max: unlimited*)

- Storage quotas are not (*presently*) strictly enforced
- Query timeouts can be extended using an API argument



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center



# Web Interfaces

## Image Cutouts

## Data Discovery

\* required field.

A list of RA and DEC in decimal degrees:

ra dec  
 159.815 -0.655  
 161.051 0.152  
 161.739 0.893  
 164.090 -0.889

Survey: NOAO Science Archive  
 Size: 0.1 degree  
 Filter:  All  u  g  r  i  z  Y

Get Image

RA:159.815\_DEC:-0.655 153 images found. [Download FITS](#)

Thumbnail	instrument_name	obs_bandpass	exptime	prodtype	proctype	date_obs	All	None
	DECam	z	158	image	InstCal	2015-03-28T23:50:39.919441	<input type="checkbox"/>	<input type="checkbox"/>

[datalab.noao.edu/tap](http://datalab.noao.edu/tap)

- allwise
- dad\_dr1
- dad\_dr2
- decaps\_dr1
- des\_dr1
  - des\_dr1.des\_allwise
  - des\_dr1.des\_gaia1
  - des\_dr1.des\_gaia2
  - des\_dr1.des\_galex
  - des\_dr1.des\_hsc2
  - des\_dr1.des\_nscdr1
  - des\_dr1.des\_simbad
  - des\_dr1.flux
  - des\_dr1.galaxies
  - des\_dr1.img2coadd
  - des\_dr1.mag
  - des\_dr1.main
  - des\_dr1.neighbors
  - des\_dr1.stars
  - des\_dr1.tile\_info
- des\_sval1
- gaia\_dr1

Please login to use VOSpace and myDB.

Column Information Query Interface

Choose a database in the left panel then select the table you want

(The bold columns are indexed columns)

Column Name	Description	
a_image	Major axis size based on an isophotal model	
alphawin_j2000	Right ascension for the object, J2000 in ICRS system (full precision but not indexed)	
awin_image_g	Major axis size, from 2nd order windowed moment measurements	
awin_image_i	Major axis size, from 2nd order windowed moment measurements	REAL
awin_image_r	Major axis size, from 2nd order windowed moment measurements	REAL
awin_image_y	Major axis size, from 2nd order windowed moment measurements	REAL
awin_image_z	Major axis size, from 2nd order windowed moment measurements	REAL

### Discover your data!

Use the data discovery tool below to explore image maps and overlay selected catalogs.

**Base Image Layers:**  
 NOAO ExposureMap  
 GAIA DR2 flux map  
 Hydrall

**Choose Overlay Image Layers: Color Map:**  
 1. None native Reverse  
 2. None native Reverse  
 3. None native Reverse  
 4. None native Reverse

**Color map:**  
 rainbow Reverse

**Overlay Catalog Layers:**  
 NOAO Image Centers  
 Smash Catalog

Column Information Query Interface

\* required field.  
 All queries must be written in ADQL.  
 To issue direct SQL database queries, use the queryClient module or datalab command-line tool.

SELECT TOP 10 \* FROM is\_dr6tractor\_primary  
 -- Show all of the columns for 10 objects in the Legacy Survey DR6 primary table

Query: clear  
 Query examples:  
 Get table columns  
 Perform a box search  
 Make an object density map  
 Extract a time series

Preview Limit: 1000 (Maximum rows to display: Not to exceed 10000)  
 Sort Column:  
 Sort Order: ascending

Process  Download results (Maximum rows to download: Not to exceed 500,000)

Results 1-10 of 10 (10 before filtering) Show 10 results per page

Text boxes under each column define filters to select rows matching the condition (e.g. <38.59.00) [Apply Filter](#) [Clear Filter](#)

[Select All Rows](#) [Unselect All Rows](#) [Show Row 2 Values](#)

decals_id	hmr1	ra	dec	elon	elat
Number	Number	Number	Number	Number	Number
6597107313347073	3618229	13:16:20.73	47:03:15.0	172.32439251049684	49.48293937928424
6597107313347080	3618229	13:16:35.32	47:03:07.4	172.37828178107118	49.50520661941243
6597107313347081	3618229	13:16:04.07	47:03:08.3	172.2668474618771	49.45407023920262
6597107313347082	3618231	13:16:52.56	47:03:10.6	172.4389843433842	49.53424493146940
6597107313347083	3618229	13:16:24.36	47:03:11.0	172.33832104400698	49.48801890399619

## Query Interface

## Schema Browser



National Optical Astronomy Observatory  
 Cerro Tololo Inter-American Observatory  
 Kitt Peak National Observatory  
 Community Science and Data Center





# API and Command-line

## Python API

```
from dl import storeClient as sc
```

```
sc.ls(format='long')
```

```
sc.ls('vos://sdss/spec')
```

```
data = sc.get ('mydata.csv')
```

```
stat = sc.put ('./mydata.csv', to='/sdss')
```

```
stat = sc.cp ('*.fits', '/raw')
```

## datalab Command Client

```
% datalab ls -l
```

```
% datalab ls vos://sdss/spec
```

```
% datalab get 'vos://*.fits' out='./raw/'
```

```
% datalab put ./mydata.csv to='vos://sdss'
```

```
% datalab mkdir '/raw'
```

```
% datalab mv fr='*.fits' to='/raw'
```



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# API and Command-line

## Python API

```
from dl import queryClient as qc  
  
query = 'select * from usno.a2'  
df = qc.query(query, fmt='pandas')  
  
jobID = qc.query (adql=query)  
while qc.status(jobID) != 'COMPLETED':  
    time.sleep(1)  
data = qc.results (jobID)
```

## datalab Command Client

```
% sql = 'select * from usno.a2'  
% datalab query $sql fmt=csv out=mydata.csv  
  
% datalab query $sql async=true out='vos://mydata'  
  
% datalab query $sql out='mydb://mydata'  
% datalab import mydata.csv mydata
```



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# Online Documentation

## User Manuals

- Overviews
- *HowTo* guides
- API documentation
- Detailed science cases
- Glossary
- SQL examples
- Tips & tricks

NOAO Data Lab 1.1.0 documentation »

### Welcome to the NOAO Data Lab documentation

Contents:

- 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

Indices and tables

- Index
- Module Index
- Search Page

NOAO Data Lab 1.1.0 documentation »

© Copyright 2016-2017, NOAO Data Lab <datalab@noao.edu>. Last updated on 2018-Jan-07.

### 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]



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center

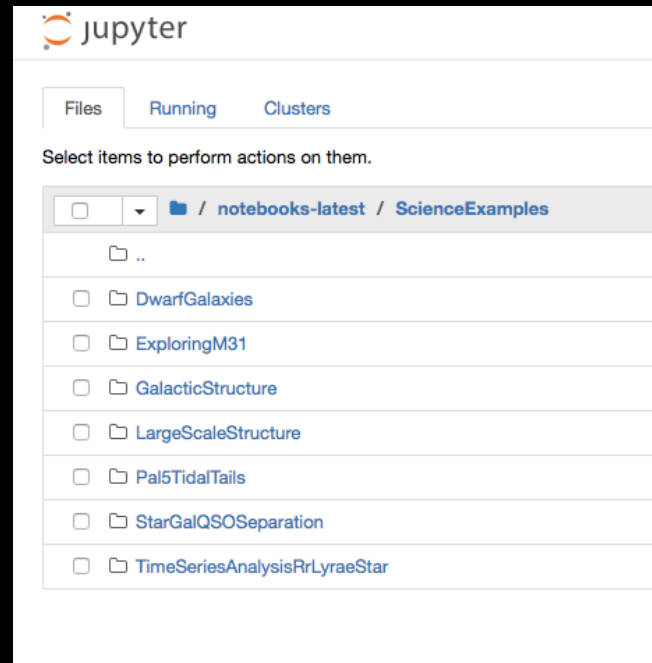




# Notebook Server

## Default Notebooks

- *Getting Started* guides
- *HowTo* guides
- Data Access Overviews
- Science Examples



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center







# Notebook Server

## Default Notebooks

- *Getting Started* guides
- *HowTo* guides
- Data Access Overviews
- Science Examples

### 5.2 Construct and Submit a SQL Query

We apply a box search, selecting the RA and Dec limits that align with the region indicated above.

```
In [67]: %%time
## Write SQL query statement as a string.
query2 = """
SELECT obj_id, ra, dec, mean_mag_g_r, mean_mag_r_i,
mean_mag_g, mean_mag_r, mean_mag_i, fracflux_g, fracflux_r, fracflux_i
FROM decaps_drl.object
WHERE (ra>134.7 AND ra<135.7 AND dec<-43.7 AND dec>-44.9
AND NOT mean_mag_g_r='Nan' AND NOT mean_mag_g_r='Infinity' AND NOT mean_mag_g_r='-Infinity'
AND NOT mean_mag_r_i='Nan' AND NOT mean_mag_r_i='Infinity' AND NOT mean_mag_r_i='-Infinity')
LIMIT 100000"""

# Submit the SQL query and return the output as a Pandas dataframe.
df2 = qc.query(sql=query2, fmt='pandas')

CPU times: user 58 ms, sys: 9 ms, total: 67 ms
Wall time: 12.4 s
```



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center



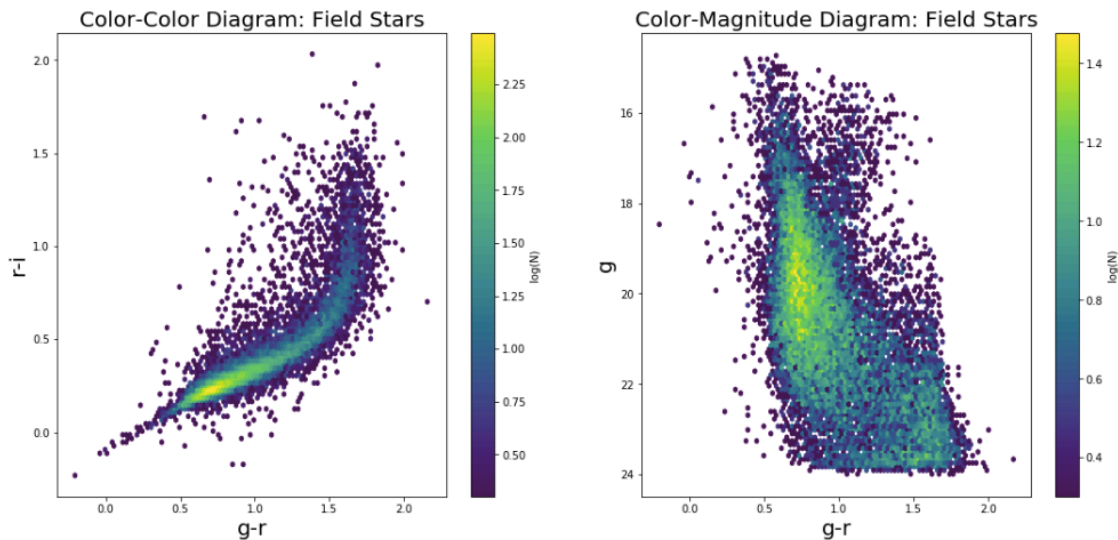


# Notebook Server

## Default Notebooks

- *Getting Started* guides
- *HowTo* guides
- Data Access Overviews
- Science Examples

```
ax2.set_ylabel(y2_label,fontsize=20)  
ax2.set_ylim([max(y2) + 0.5, min(y2) - 0.5])  
cb2 = plt.colorbar(im2,ax=ax2,label='log(N)')  
  
plot_results(g_r, r_i, 'Color-Color Diagram: Field Stars', 'g-r', 'r-i',  
            g_r, g, 'Color-Magnitude Diagram: Field Stars', 'g-r', 'g')
```



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center

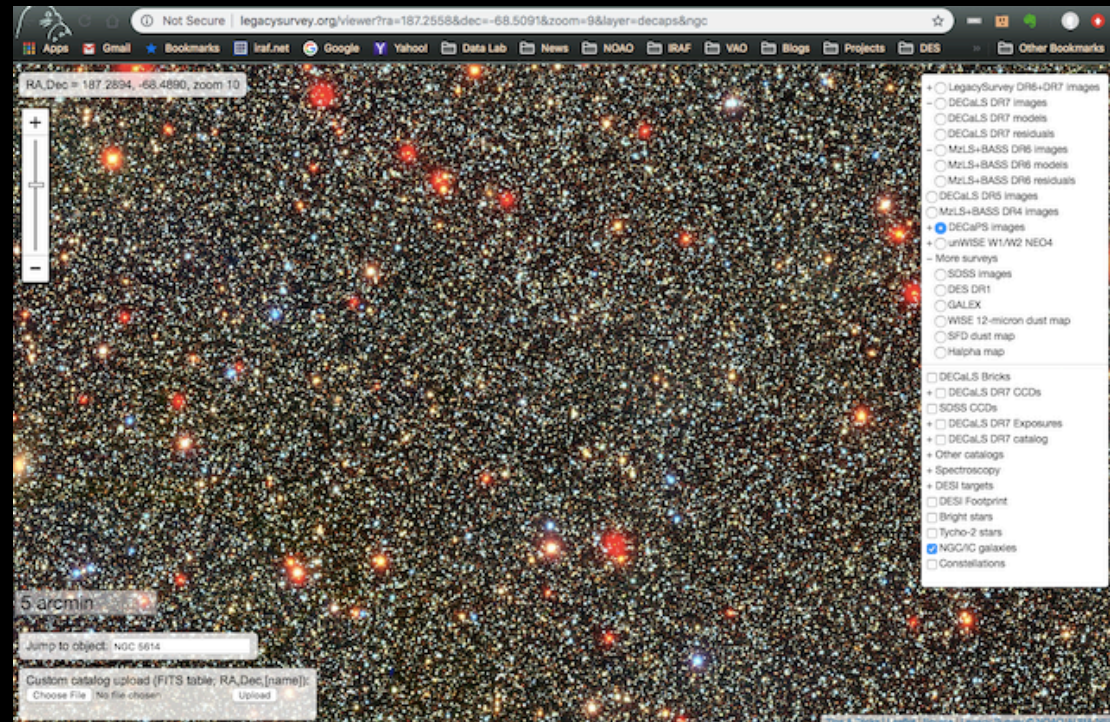




# Notebook Server

## Default Notebooks

- *Getting Started* guides
- *HowTo* guides
- Data Access Overviews
- Science Examples



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





# Community Development

## Science/Example/Tutorial Notebooks:

- Presently a “*Contact Us*” model to contribute, would like to move to a “*Self Publish*” with a directory

## Source Code:

- Repos to be made public once we complete packaging/docs and remove configuration details
- Client pull requests being accepted
- (Future) User-contributed compute containers (e.g. Docker hub)

## User Feedback!

Come see us at AAS 233 in Seattle



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center





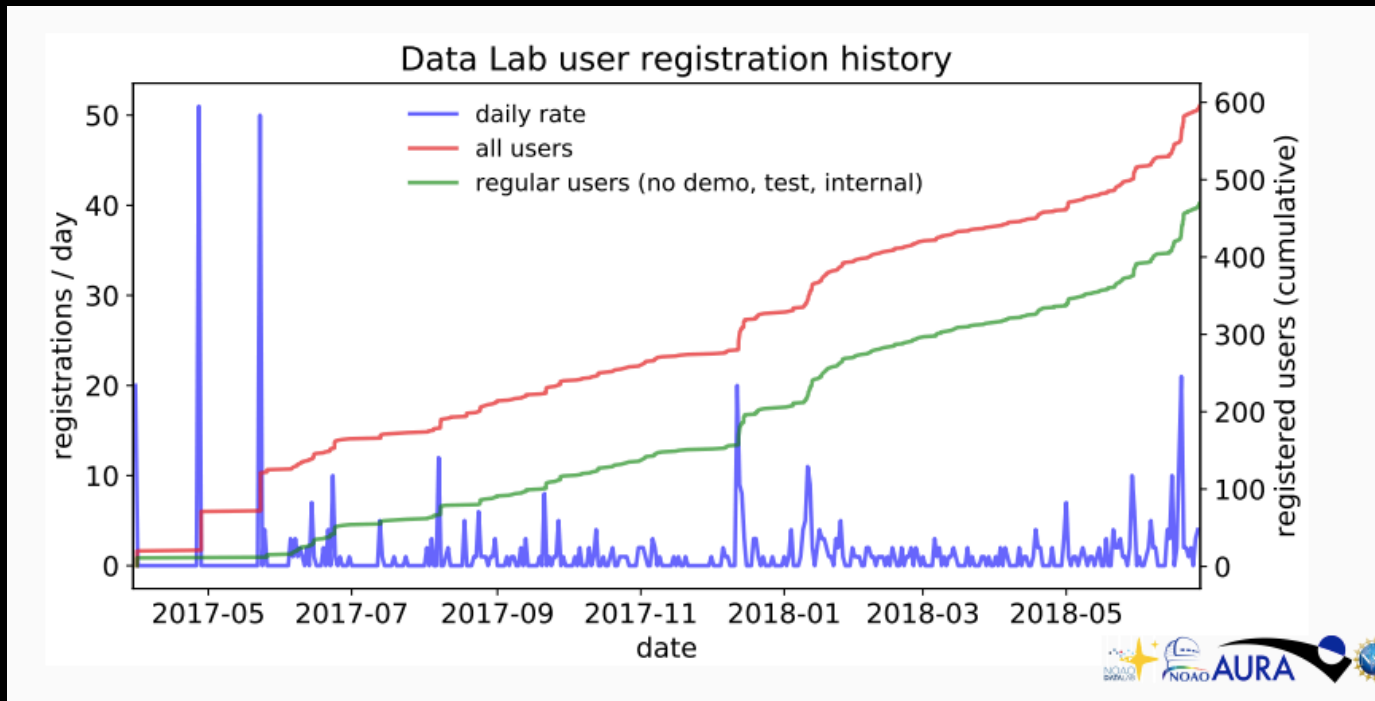
## Lessons Learned



National Optical Astronomy Observatory  
Cerro Tololo Inter-American Observatory  
Kitt Peak National Observatory  
Community Science and Data Center









# Thank You

Login | Sign up

Service Status:

About Quick Start Tools Survey Data Docs/Help Events

Get your account:

<http://datalab.noao.edu>

Contact us:

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

[@NOAODataLab](#)

Crowd-sourced survey of the sky: total exposure time (log scale) for all cameras during 2004-2017 (log scale)



The NOAO Data Lab is operated by the National Optical Astronomy Observatory, the national center for ground-based nighttime astronomy in the United States operated by



