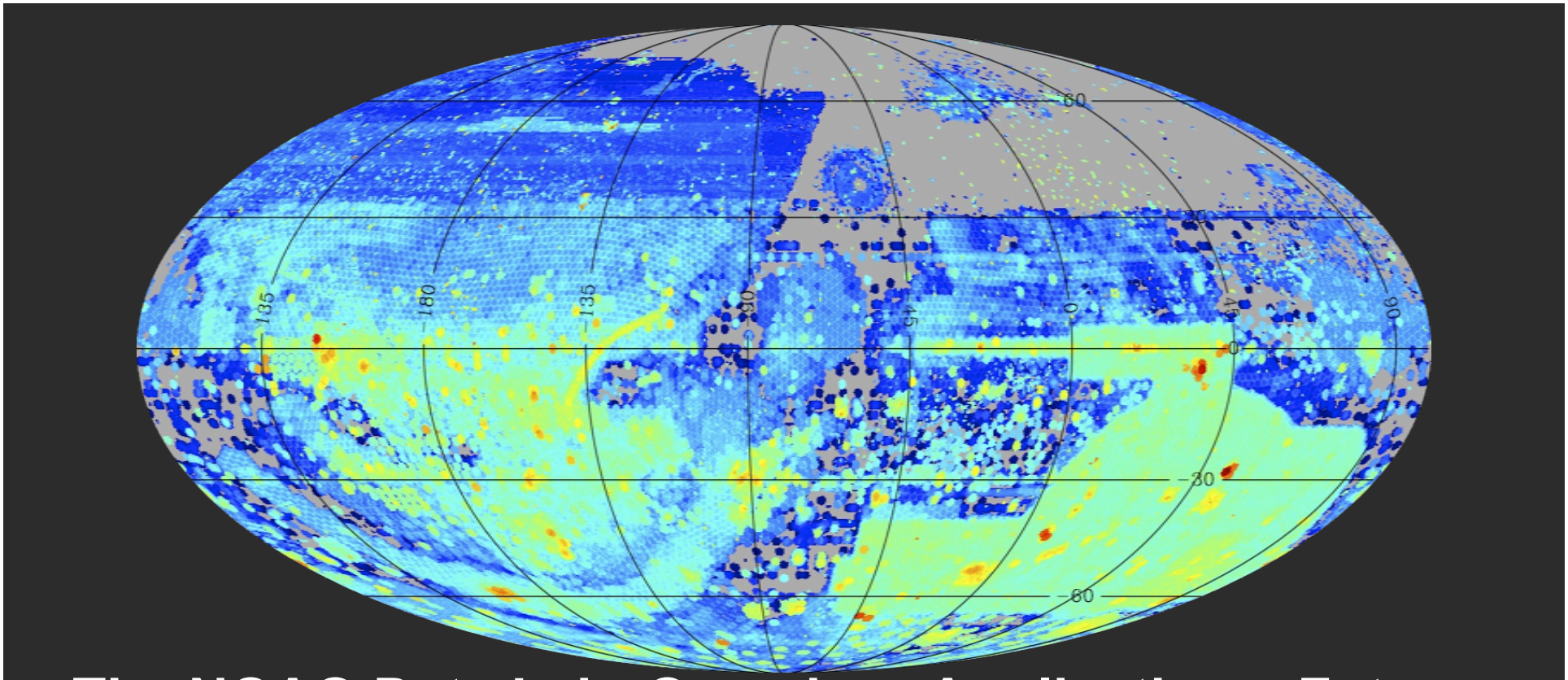


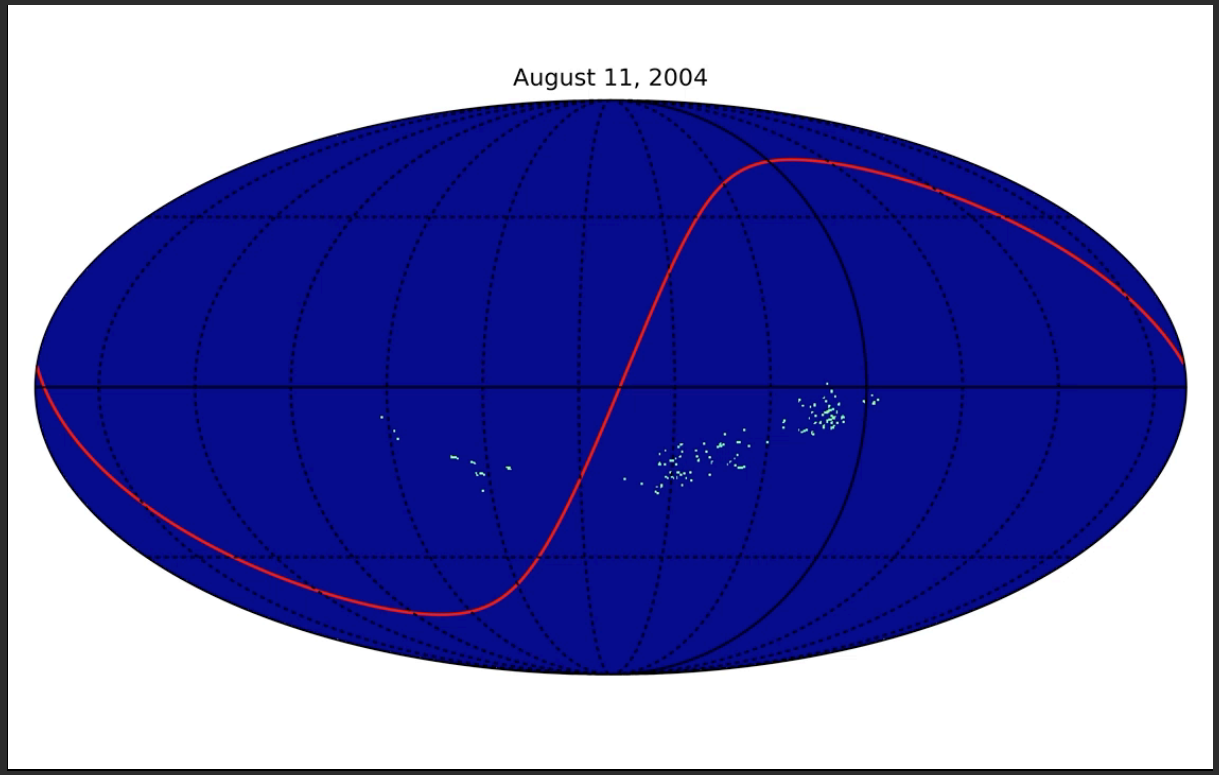
The NOAO Data Lab: Overview, Applications, Future
What's So Great About the Data Lab is…

Leah Fulmer, on behalf of the Data Lab Team

We acknowledge that we are currently situated on the traditional lands of the Tohono O’odham Nation.

Furthermore, we would like to thank the cleaning and maintenance staff of this building and all telescope facilities, without whose labor this work could not be conducted.

Motivation



Sky coverage from NOAO wide-field imaging cameras on the 4m telescopes

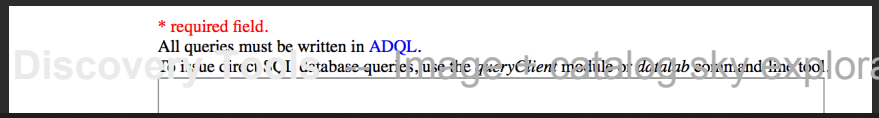
Mission

The NOAO Data Lab seeks to **empower astronomers** through **efficient exploration and analysis** of large astronomy datasets with an **emphasis on NOAO wide-field 4m telescopes**.

Its primary objectives are to **connect users with high-value catalogs** through both user-defined and data discovery search forums, to **build Python-based tools** for the efficient visualization and analysis of data both within and beyond the Data Lab system, and to **provide direct service** to astronomers in order to optimize user experience.

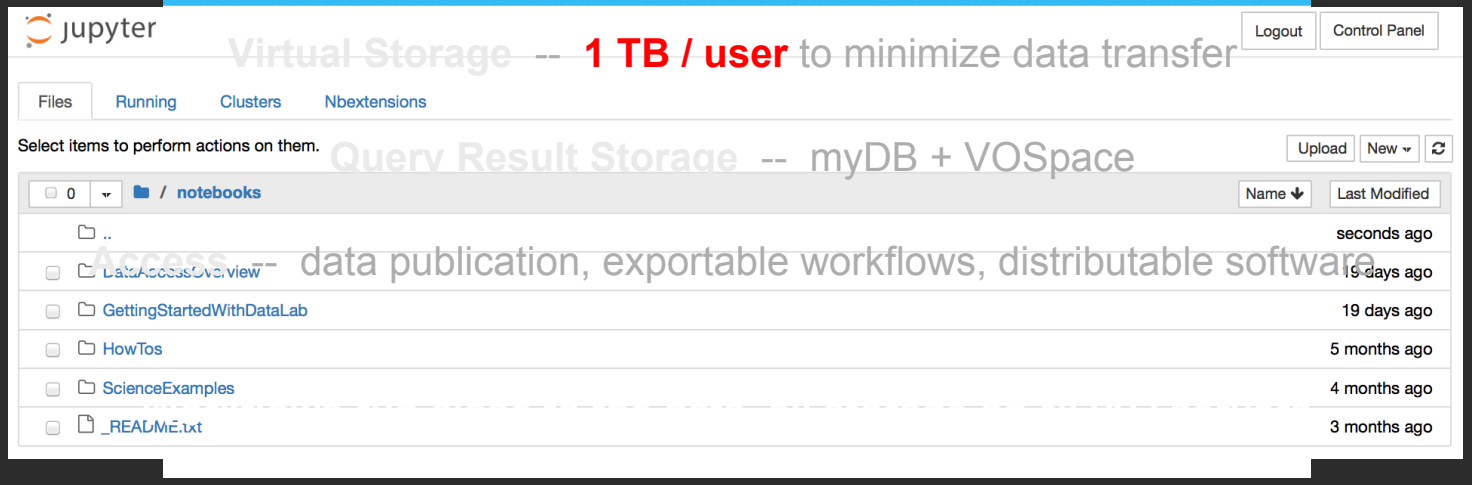
Approach

Catalog + Image Query -- Python queryClient



Data Discovery -- Image + catalog sky exploration

Analysis Jupyter Notebook server + scripted analysis

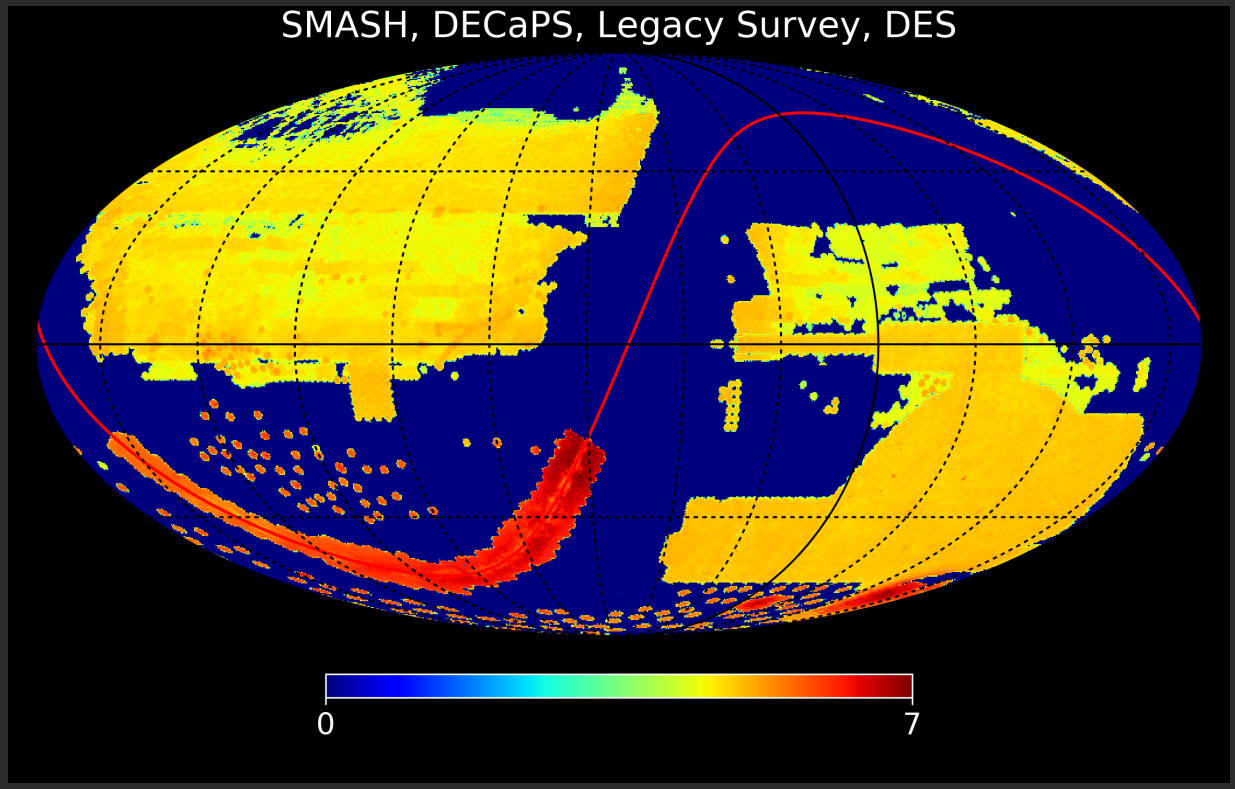


Virtual Storage -- **1 TB / user** to minimize data transfer

Query Result Storage -- myDB + VOSpace

Access -- data publication, exportable workflows, distributable software

Catalog Exploration

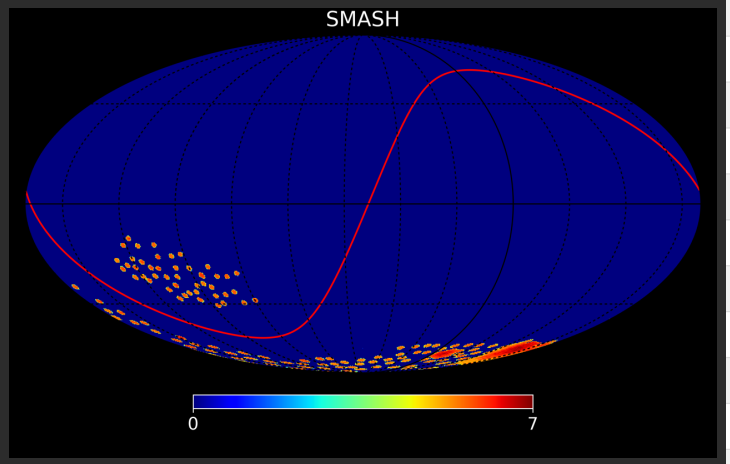


Catalog Exploration

Survey of the Magellanic Stellar History

Intended to identify **distributed, low surface brightness stellar populations** associated with the stellar halos of **tidal debris** from the Magellanic Clouds.

Survey **in-progress**, having observed ~100 million objects over an expected 250 million objects



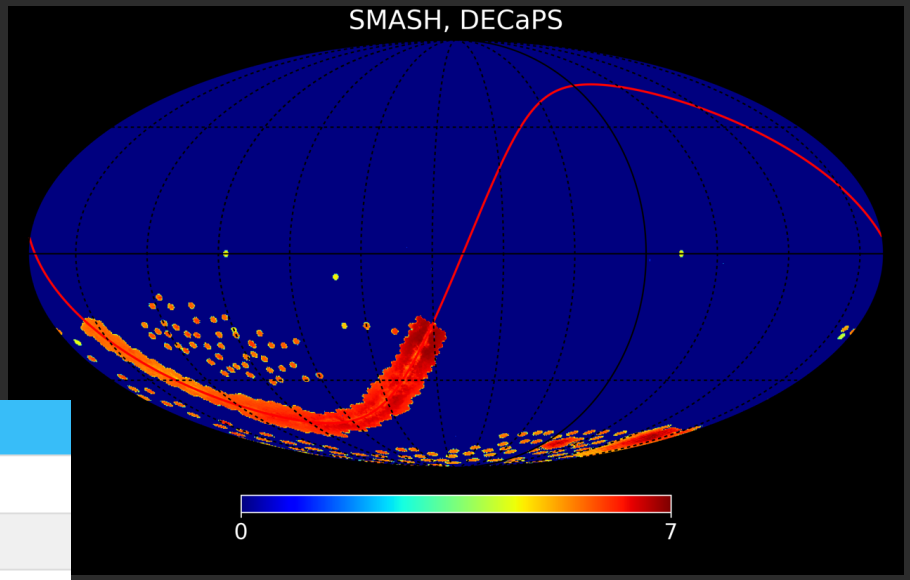
Shout out: **David Nidever** at **2:40pm**

SMASH at a Glance	
Area covered	480 deg^2 spanning ~2400 deg^2
Bands	*ugriz*
Depth (5σ, *ugriz*)	23.9, 24.8, 24.5, 24.2, 23.5 mag
Seeing (*ugriz*)	1.22, 1.13, 1.01, 0.95, 0.90 arcsec
Number of fields	197
- DR1	61
Number of DECam exposures	5,809
- DR1	2,480
Number of objects	~420,000,000
- DR1	101,425,210
Number of measurements	~4,000,000,000
- DR1	722,653,189
Photometric precision	~1% in *u* and 0.5-0.7% in *griz*
Photometric calibraton accuracy	~1.3% in all bands
Astrometric accuracy	~20 mas

Catalog Exploration

DECam Plane Survey

- **Multi-band imaging survey** of the Galactic Plane (grizY)
- Explore the **entire DECaPS band-merged catalog** with the Data Lab, as well as image header tables,zero points, flat fields, etc.



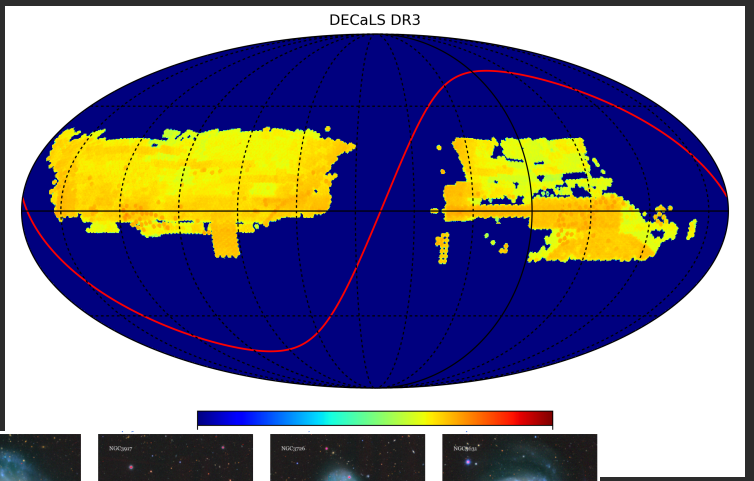
DECaPS Summary Table	
Area covered	~1000 deg^2
Bands	*grizY*
Depth (5σ, *grizY*)	23.7, 22.8, 22.2, 21.8, 21.0 mag
Number of objects	~2,000,000,000

- Stay tuned for a **science example!**
- Shout out: **Eddie Schlafly** at **1:30pm**

Catalog Exploration

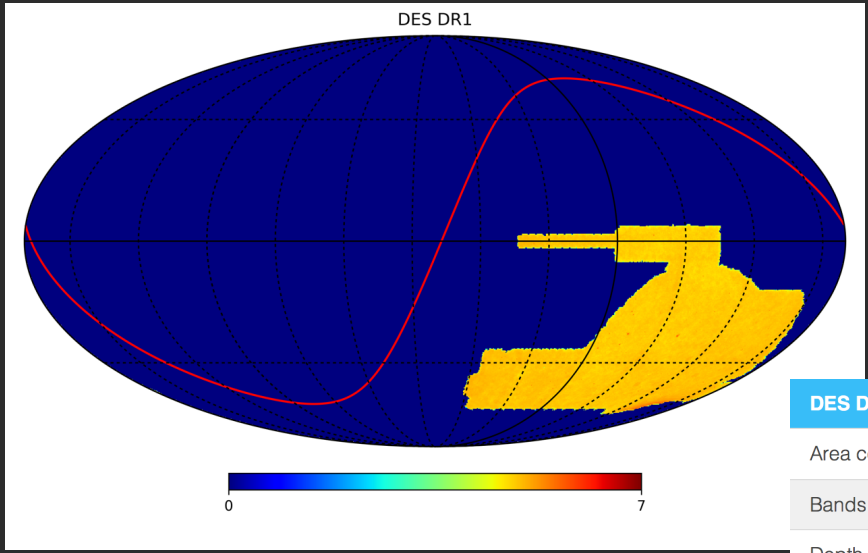
DECam Legacy Survey

- **High-quality optical imaging (g,r,z)** covering two-thirds of the Dark Energy Spectroscopic Instrument (DESI) footprint, ~ 9500 deg^2
- **Depths:** g = 24.7, r = 23.9, z = 23.0
- Explore **images and catalogs** with Data Lab
- **Key component of the DESI Imaging Legacy Survey**, an optical + infrared survey of the entire DESI footprint, to be utilized in target selection.
- Shout out: **Martin Landriau** at **10:40am.**



Catalog Exploration

Dark Energy Survey



- Data Lab releases include **crossmatch tables** with GALEX, VISTA Hemisphere **Survey,** WISE, and [very soon] Gaia DR2
- Search catalog of **neighboring objects** within 30 arcsec

Explore catalog of 300 million distant galaxies and 100 million Milky Way stars – **stay tuned for a science example!**

DES DR1 Summary	
Area covered	5000 deg^2
Bands	*grizY*
Depth (10σ, *grizY*)	24.45, 24.3, 23.5, 22.90, 21.70 mag
Seeing (*grizY*)	~1 arcsec
Number of Tiles	10338
Number of objects	399,263,026

Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

Setting the scene…

- An imaging survey includes a **zoo of different astronomical objects**. There are **foreground stars** from our own Milky Way as well as **background galaxies** at various distances, including **QSOs** with actively accreting supermassive black holes.
- The DECaLS Catalog **classifies these objects** through a series of morphological parameters.

Therefore…

- In the following example, we will **examine the DECaLS object classifications** within the Data Lab Environment.

We hope to demonstrate…

- Easy access of DECaLS through the **Query Manager**.
- Basic classification quality exploration through a **Jupyter Notebook Server**.

Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

In [1]:
```
__author__ = 'Stephanie Juneau, NOAO Data Lab Team'
__version__ = '20180104' # yyyymmdd
__datasets__ = ['ls_dr3']
```

Star/Galaxy/QSO Classification in the DESI Imaging Legacy Surveys

by Stéphanie Juneau, Robert Nikutta, Knut Olsen and the NOAO Data Lab Team

In this notebook, we investigate the optical and infrared colors of astronomical sources detected in the DECam Legacy Survey (DECaLS). The third data release of this imaging survey comprises ~400 millions stars, galaxies and quasars (or QSOs: Quasi-Stellar Objects).

NOAO Data Lab products and services used here:

- the Legacy Survey (LS) DR3 database
- Jupyter Notebook Server
- Query Manager
- Image cutout tool similar to Data Lab SIA

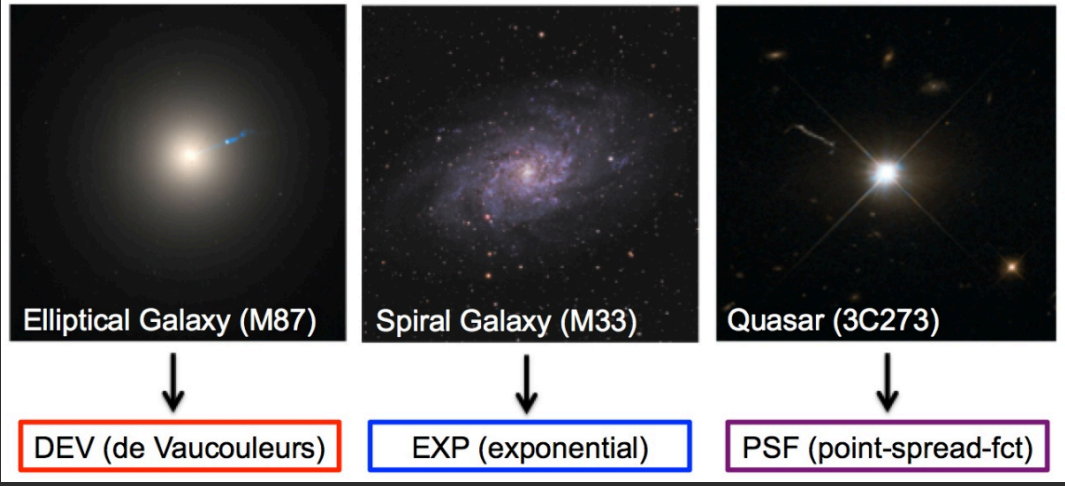
Below, we query the database, compute colors, plot a few color combinations, and take into account the source "type" as defined from the light profile shape in order to differentiate between object classes.

Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

Possible shapes for DECaLS DR3:

- PSF (point spread function: size will vary with the seeing of the observations)
- SIMP ("simple" galaxies: round, exponential profile with 0.45″ effective radius)
- EXP (exponential profile; spiral galaxies)
- DEV (deVaucouleurs profile; elliptical galaxies)
- COMP (composite deVaucouleurs+exponential at same centroid)



Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

Step 1: Construct and Submit a SQL Query

Wavebands: g,r,z from DECam **Morphological parameter:** type **Data quality:** S/N in each band

```
# Write query statement (sql) as a string
# NOTE: triple quotes allow us to break the string on multiple lines
query = """
        SELECT dered_mag_g as gmag, dered_mag_r as rmag, dered_mag_z as zmag,
               type,
               snr_g, snr_r, snr_z, ra, dec
        FROM ls_dr3.tractor_primary
        WHERE (snr_g>3 and snr_r>3 and snr_z>3)
        LIMIT 400000"""

# dered_mag_g,r,z  = AB magnitudes in DECam g,r,z bands corrected for Galactic reddening
# dered_mag_w1,w2  = AB magnitudes in WISE bands W1 & W2 corrected for Galactic reddening
# type             = object type (PSF, SIMP, EXP, DEV, COMP)
# snr_g,r,z        = pre-computed signal-to-noise ratios (S/N) in g,r,z bands
# ra,dec           = celestial coordinates
#
# WHERE: requirement that S/N>3 in each DECaLS band
# LIMIT: returns 400,000 rows that satisfy the query
```

Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

Step 2: Format Query Output

Convert output from query to pandas DataFrame

```
# Reformat output into a table
#result = Table.read(StringIO(response), format='csv')  #dictionary
# Reformat output into a Pandas Data Frame
result = helpers.convert(response,'pandas')

# Print a few rows from the result table
print(result[:5])
print(" ")
print('This query contains', len(result), 'objects.')
```

```
Returning Pandas dataframe
      gmag     rmag     zmag  type      snr_g      snr_r      snr_z  \
0  19.2910  18.6569  17.9649  SIMP  253.79800  286.54200  295.84000
1  21.7634  20.2850  19.3403   PSF   45.30160  117.72700  121.92700
2  22.1439  20.9485  19.8895  SIMP   27.34160   56.48310   57.96450
3  23.3501  22.8142  22.2988  SIMP    9.26451   11.25470    6.50289
4  23.7298  23.1004  21.6915  SIMP    6.59172    8.75818   11.35560

          ra        dec
0  42.551770 -19.478720
1  42.542098 -19.483807
2  42.552989 -19.488538
3  42.555074 -19.487900
4  42.551644 -19.473996

This query contains 400000 objects.
```

Inspect output data

Check object limit

Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

Step 3: Create Color Variables Define Quality Threshold

Define colors with S/N quality threshold

```
# Select range of interest
thres = 5.   #threshold value for S/N (here, making it more stringent than query)
keep = (result['snr_g']>thres)&(result['snr_r']>thres)&(result['snr_z']>thres)

# Colors
g_r   = result['gmag'][keep] - result['rmag'][keep]
r_z   = result['rmag'][keep] - result['zmag'][keep]

# Classification per object type
objtype = result['type'][keep]

print('After enforcing a S/N threshold, this query now contains', len(objtype), 'objects.')
```

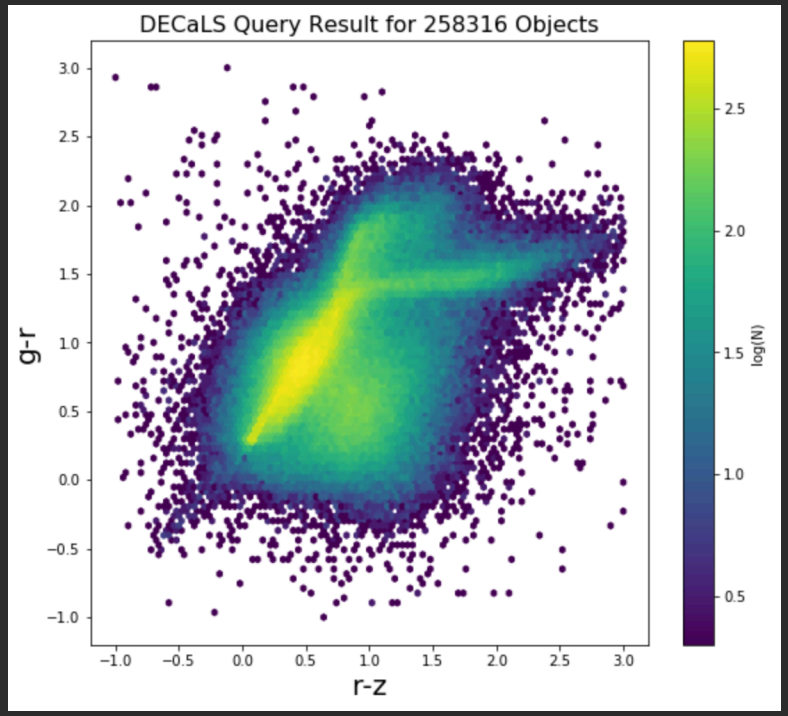
```
After enforcing a S/N threshold, this query now contains 258316 objects.
```

Check new output objects

Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

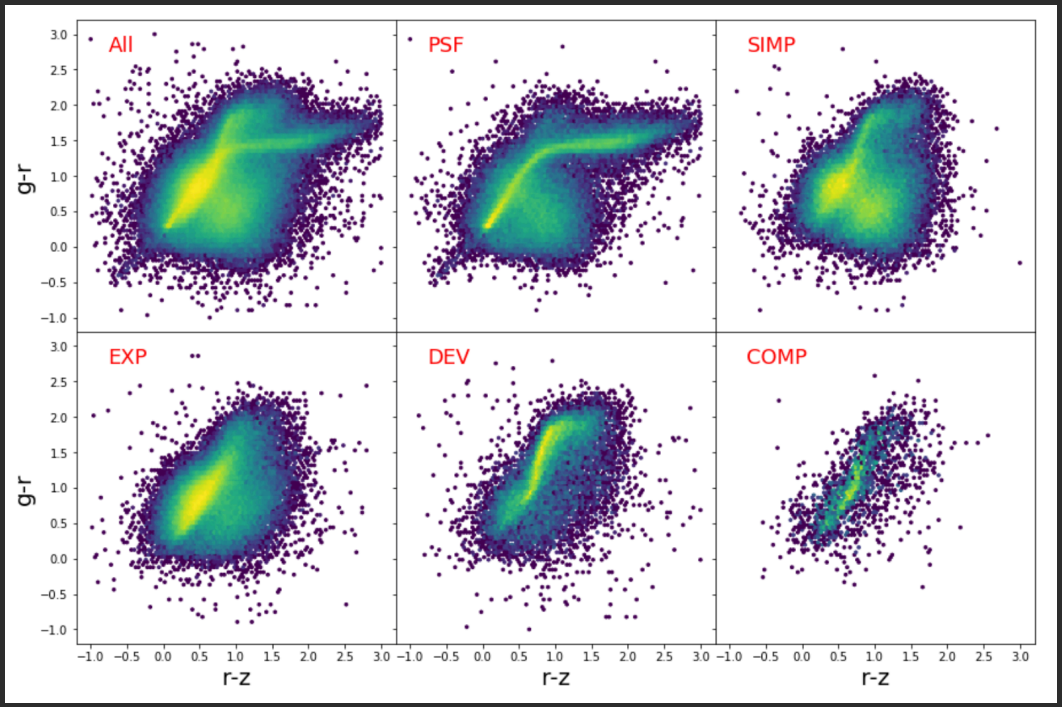
Step 4: Plot Initial Results



Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

Step 5: Separate Objects by Classification



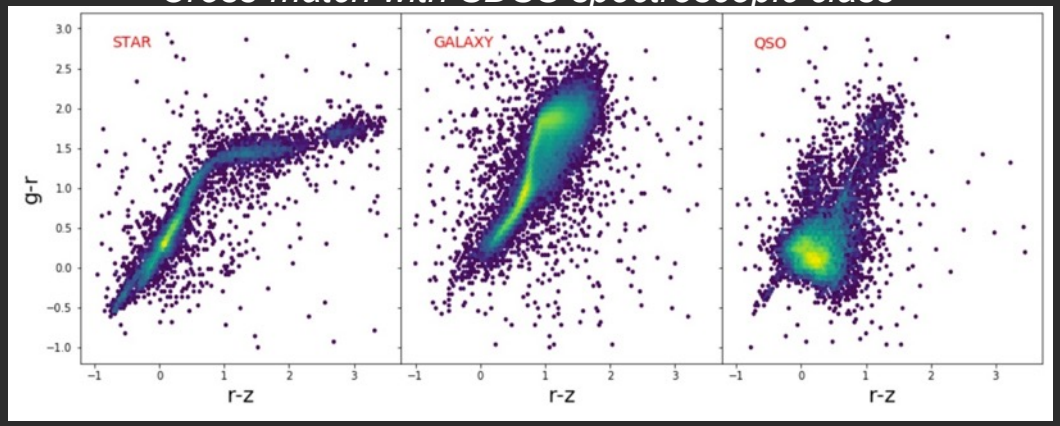
Star / Galaxy Classification

Exploring DECaLS within the Data Lab Environment

Following Steps

Joint Query

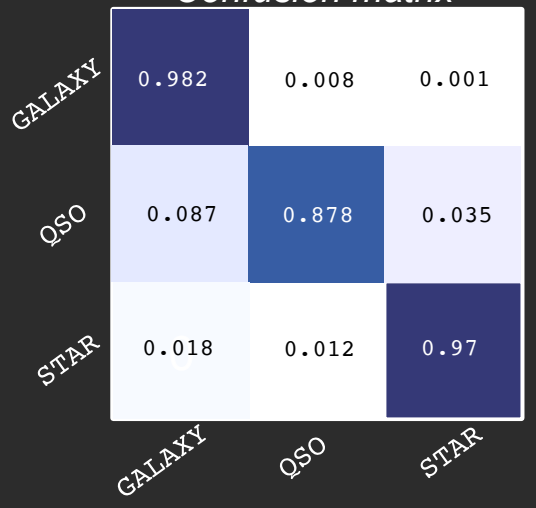
Cross-match with SDSS spectroscopic class



Modified version of figure by Bela Abolfathi (UC Irvine)

Machine Learning

Confusion matrix



Courtesy of Jan-Torge Schindler (Univ. of Arizona)

Galactic Structure

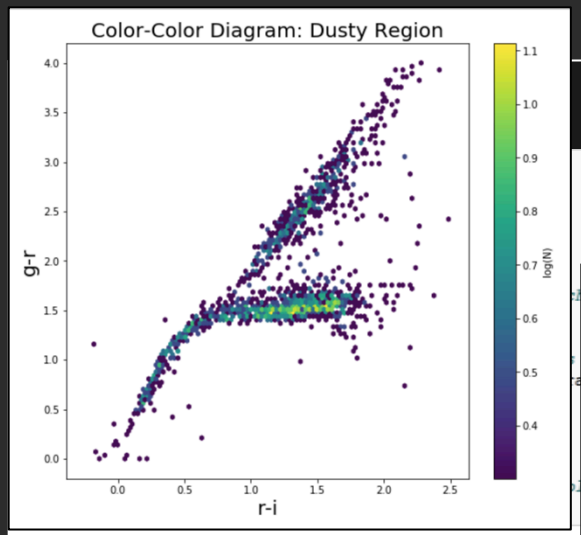
Exploring DECaPS with the Data Lab

The DECam Plane Survey

Construct and Submit a SQL Query

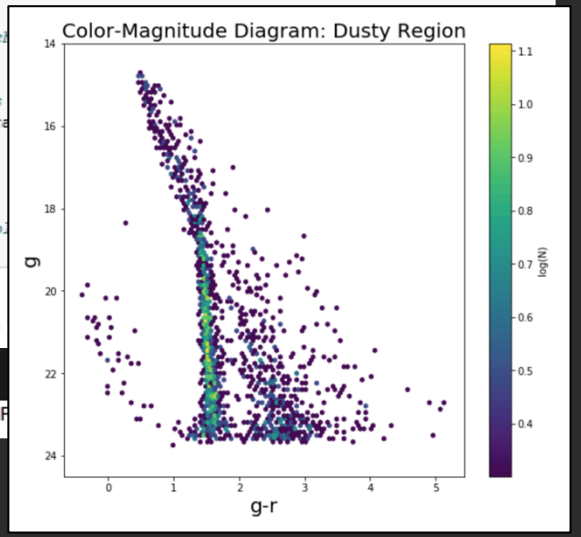
Format Query Output and Define Quality Threshold

Plot Results




```
Returning Pandas dataframe
This query initially contained 2151 objects.
After data quality cuts, this query now contains 1819
```

DECaPS Galactic Plane as shown in the Legacy Survey Sky Viewer. F



Preparing for DESI
Overview

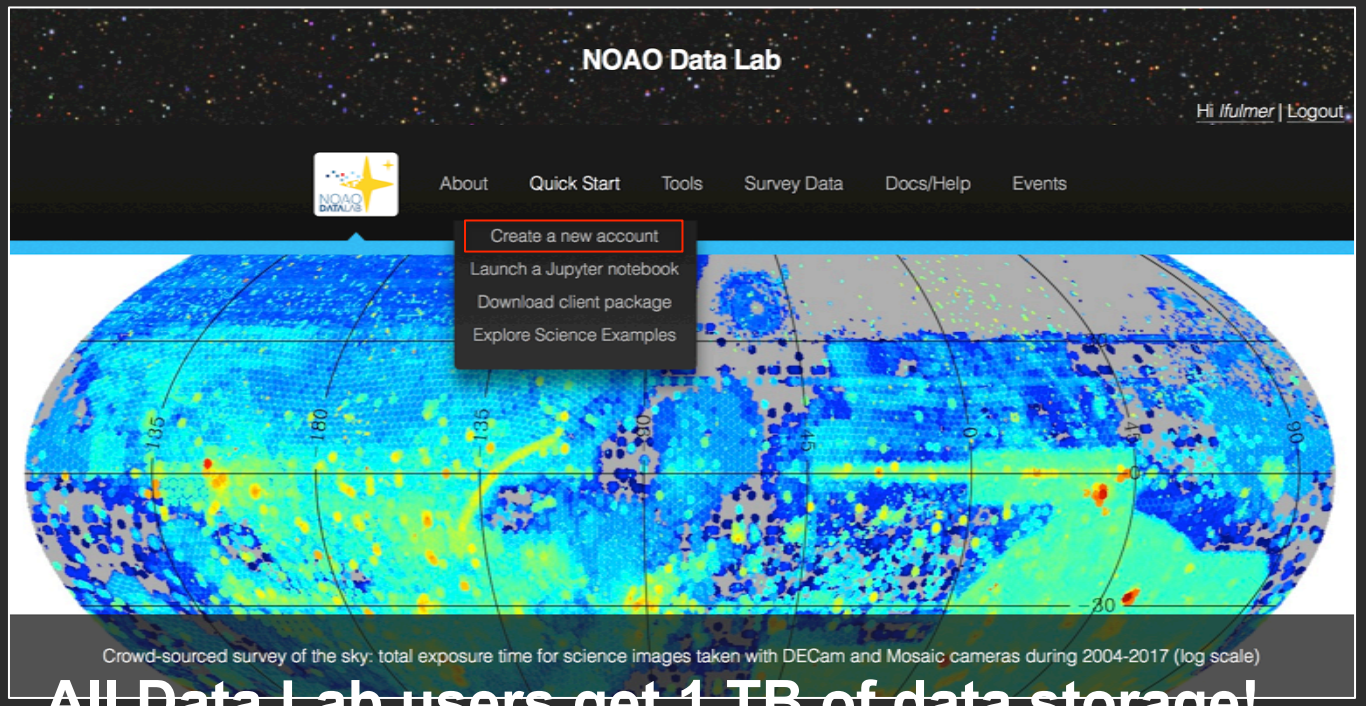
- **14,000** deg^2
- **10 million** spectra of stars
- **30 million** spectra of galaxies and quasars
- Commissioning starts in **2018**
- Survey runs **2019-2024**

Object Class	Number of Spectra	Redshift Range
bright galaxies, r < 19.5	10 million	0 < z < 0.4
luminous red galaxies (LRGs)	4.2 million	0.4 < z < 1.0
emission line galaxies (ELGs)	18 million	0.6 < z < 1.6
quasars (QSOs)	2.4 million	0.5 < z < 3.5
Milky Way stars	10 million	---

Preparing for DESI
Implications

- Up until this point, the Data Lab has **primarily served imaging data**
- With the advent of DESI, we look to provide not only **access to spectroscopic data**, but also **software tools with which to analyze spectra**
- We are currently exploring tools for **optimized functionality with DESI + generalized spectra**
- We are in close contact with the developers of **SpecViz, MOSViz, Glue,** and **Inspector,** and we welcome any suggestions or feedback as to the best path forward

Getting Started



All Data Lab users get 1 TB of data storage!

Connecting at the DECam Community Workshop

Data Lab:

- **E-mail:** datalab@noao.edu
- **Twitter:** @NOAODataLab

Myself:

- **E-mail:** lfulmer@noao.edu
- **Twitter:** @leahmfulmer

Data Lab Team:

Knut Olsen (Team Leader) Stephanie Juneau (Project Scientist) David Nidever (Data Scientist) Robert Nikutta (Data Scientist) Mike Fitzpatrick (Lead Developer) Wendy Huang (Software Engineer) Adam Scott (Database Architect) Glenn Eychaner (Software Engineer) Ben Weaver (Developer) Leah Fulmer (Developer) Steve Ridgway (Scientist)