

The Rubin Science Platform and data access - early data previews and releases

Melissa Graham Lead Community Scientist for Rubin Observatory

NAM 2022









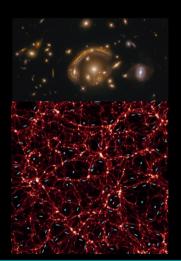




Open Astrophysical Questions

Cosmology

Understand dark energy and dark matter, and the origin and fate of the universe, by studying gravitational lensing and large-scale structures across cosmic time.



Transient Phenomena

Understand evolutionary processes by studying how stars and compact objects (e.g., black holes) change brightness, interact, merge, and explode.



The Milky Way

Understand the structure and evolution of our Galaxy's bulge, disk, and halo – and its satellites and tidal streams – by mapping the stars of the Milky Way.

The Solar System

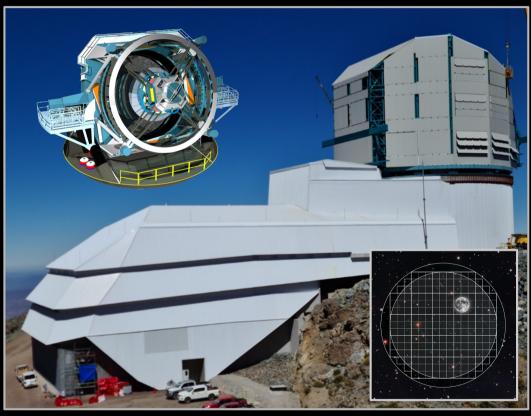
Understand the formation and evolution of our Solar System, and the risk of potentially hazardous asteroids, by making a full inventory of objects down to ~100 m scales.



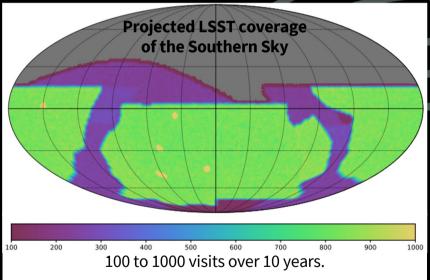




VERA C. RUBIN Legacy Survey of Space and Time (LSST)



The Vera C. Rubin Observatory, currently under construction in Chile, will reach these nextgeneration science goals by using its 8.4 m primary mirror, 9.6 deg² field-of-view camera, and six optical-NIR filters to execute the 10-year LSST.





VERA C. RUBIN Legacy Survey of Space and Time (LSST)

Billions of stars and galaxies.

Millions of transients, variables, and moving objects.

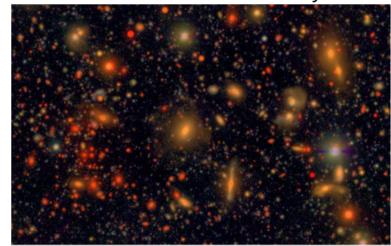
A data set of unprecedented volume and complexity.





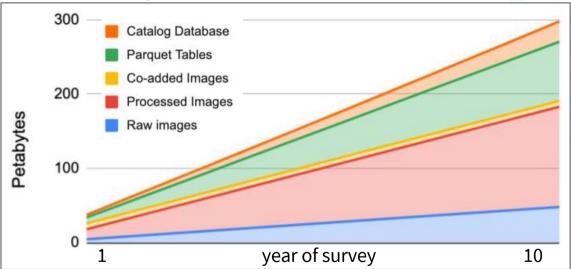
Legacy Survey of Space and Time (LSST)

The final LSST 10-year sky map will be like having ~3 million of these, tiled over the entire southern sky.



Ivezić et al. 2019

The Rubin Observatory's total data holdings will start at ~40 PB and grow to ~300 PB over the 10-year LSST.



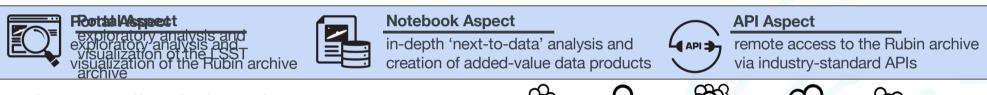
O'Mullane et al. 2021 (RTN-003.lsst.io)



Rubin Science Platform (RSP)

It will not be possible to download the entire LSST data set, and scientists will need a venue for "next-to-the-data analysis".

The **Rubin Science Platform (RSP)** is a set of integrated web-based applications and services running at the Rubin Observatory Data Access Centers (DACs).



The RSP will include tools to query, visualize, subset, and analyze the full LSST data archives in a stable software environment located "next-to-the-data", along with storage space, compute resources, and remote access options.





Rubin Data Policy

Rubin Data Policy: ls-st/rdo-013

List of international data rights holders: lst-org/scientists/international-drh-list

Data rights holders: scientists (and students) working (or enrolled) at US or Chilean institutes, and named members of the international in-kind contribution teams.

Proprietary: data products (images and catalogs) for the first 2 years after release, and access to the Rubin Science Platform.

Public: alert packets, post-proprietary data products, and the documentation and support resources such as the Rubin Community Forum (<u>Community.lsst.org</u>).



The RSP User Experience

Goal: Democratize science by removing barriers to participation in LSST.

- abundant, discoverable documentation of the end-to-end system
- clear entry points and tutorials from beginner through advanced levels
- asynchronous, distributed, friendly support
- a stable software environment with compute resources
- prioritize research inclusion and seed expertise across the community
- enable anyone to become power user and push the cutting edge with LSST



Status: In regular use by hundreds of Rubin staff and community scientists and students participating in Data Preview 0.

- deployed at 12 sites globally (e.g., data facilities, the summit)
- primary community site is the Interim Data Facility (Google Cloud)
 - ~300 users from the science community since June 2021
 - ~300 more new users are joining now (summer 2022)
 - Chile/USA/International ~ 15/45/40
- thousands of users by the start of Rubin Operations

The Rubin Science Platform was deployed at the Interim Data Facility (Google Cloud) primarily to enable Data Preview 0, which serves simulated LSST-like data. It has been terrific for development, integration, and community engagement.

9



Data Previews 0, 1, and 2

Goals of the Data Previews

- enable the community to prepare for early LSST science with the RSP
- test integration of the LSST Science Pipelines and the RSP
- use feedback on data products and RSP functionality to inform future development

Data Previews' Data Sets

- DP0: simulated LSST-like images from the DESC's Data Challenge 2 (DC2)
 - DP0.1: DC2 as processed by the DESC available in the RSP (June 2021)
 - DP0.2: DC2 as reprocessed by Rubin Data Production available in the RSP (June 2022)
- DP1: data from the commissioning camera ("comcam", single-raft; May 2024)
- DP2: commissioning data from the LSST Science Camera (March 2025)

The exact commissioning surveys remain To Be Confirmed.

Operations start expected October 2024 (first data release, DR1, would be late 2025).



Portal

Notebooks

APIs

Documentation

Support

Community

DP0.2 HiPS Coverage Maps now available in the Portal 🎉

Show more

Rubin Science Platform

Portal

Discover data in the browser



Learn more about the portal.

Notebooks

Process and analyze LSST data with Jupyter notebooks in the cloud



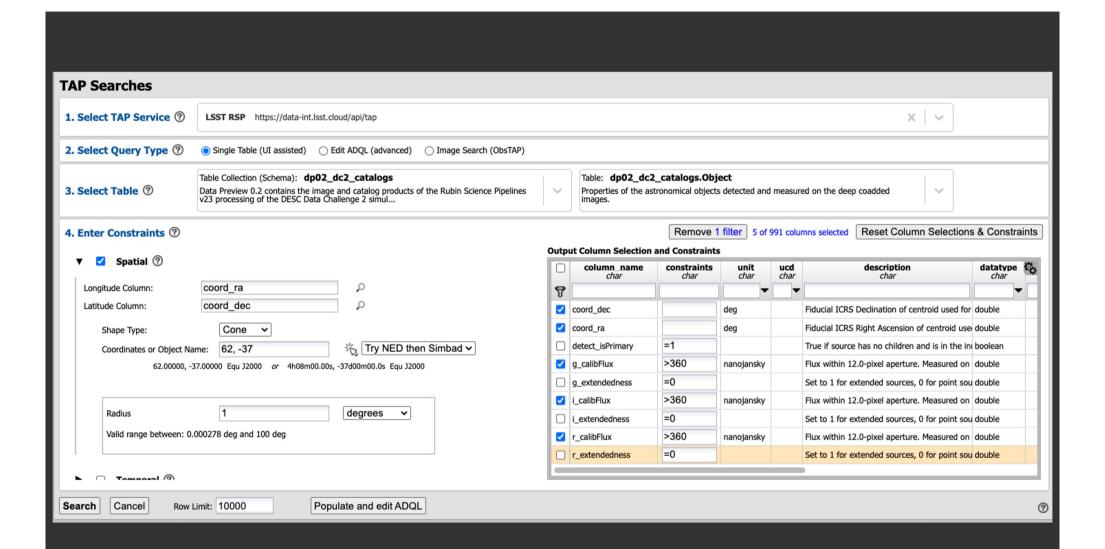
Learn more about notebooks.

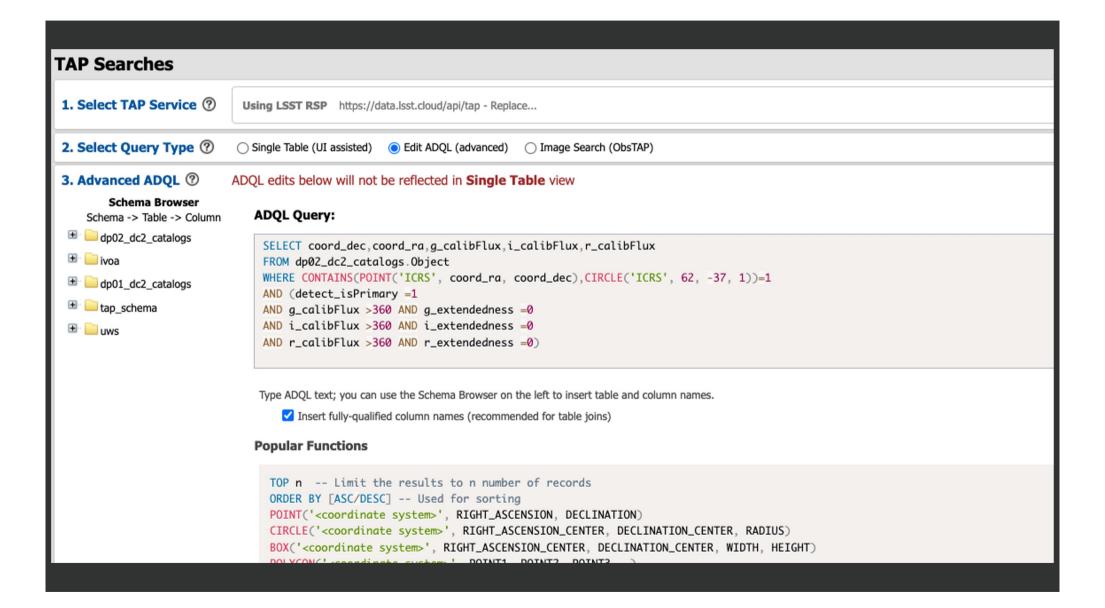
APIs

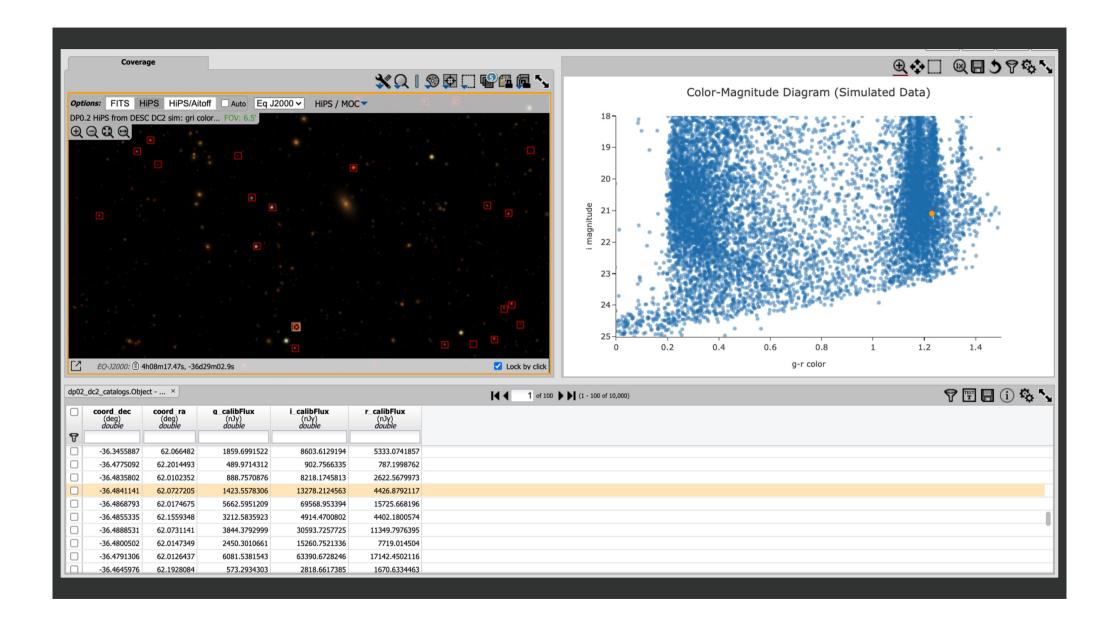
Learn how to programatically access data with Virtual Observatory interfaces

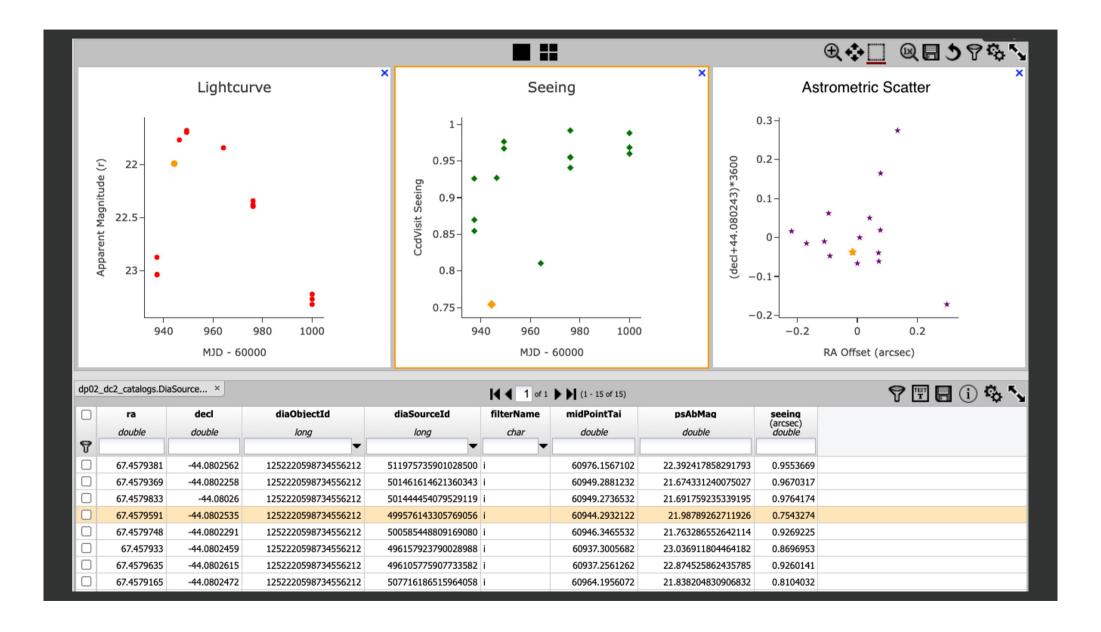
Log in

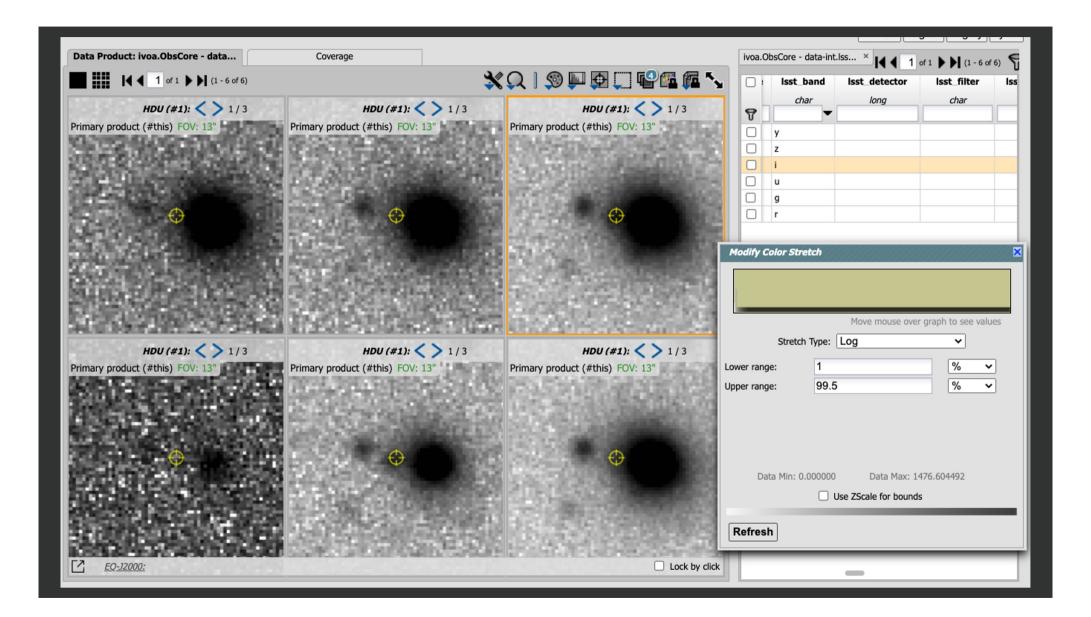














Portal

Notebooks

APIs

Documentation

Support

Community

DP0.2 HiPS Coverage Maps now available in the Portal 🎉

Show more

Rubin Science Platform

Portal

Discover data in the browser



Learn more about the portal.

Notebooks

Process and analyze LSST data with Jupyter notebooks in the cloud



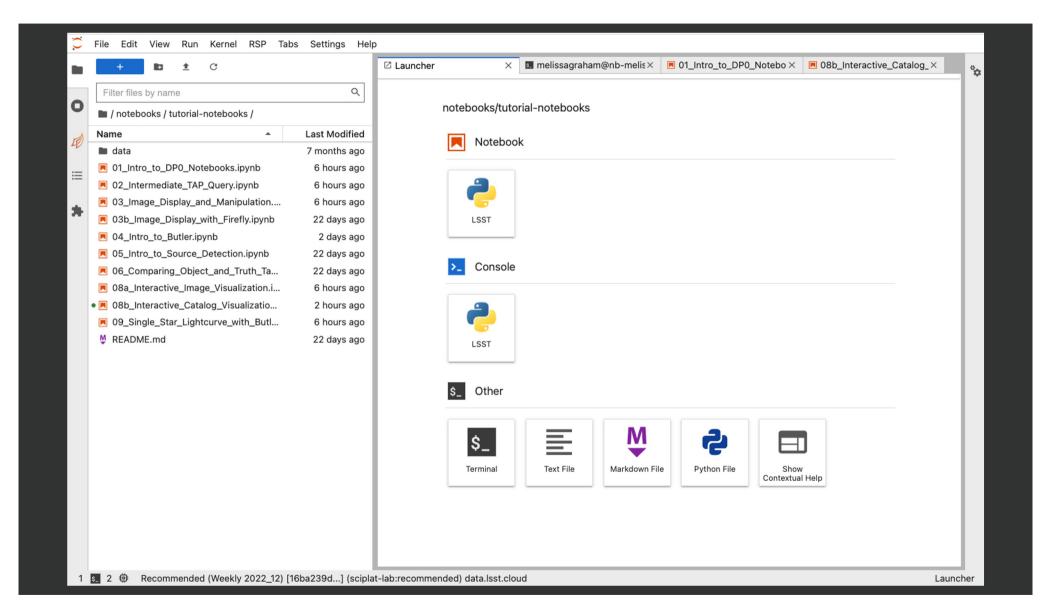
Learn more about notebooks.

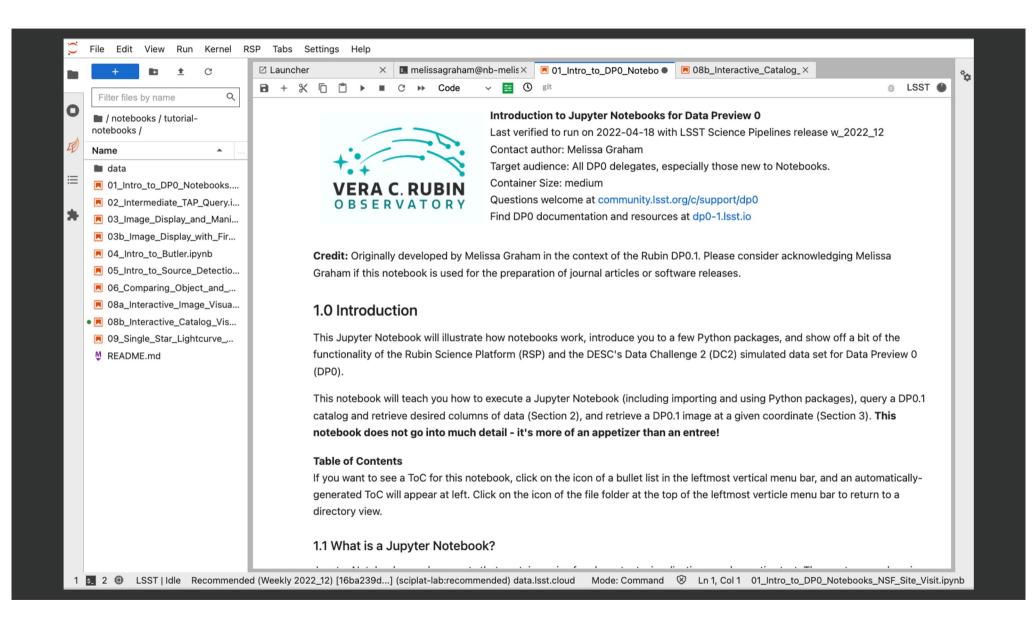
APIs

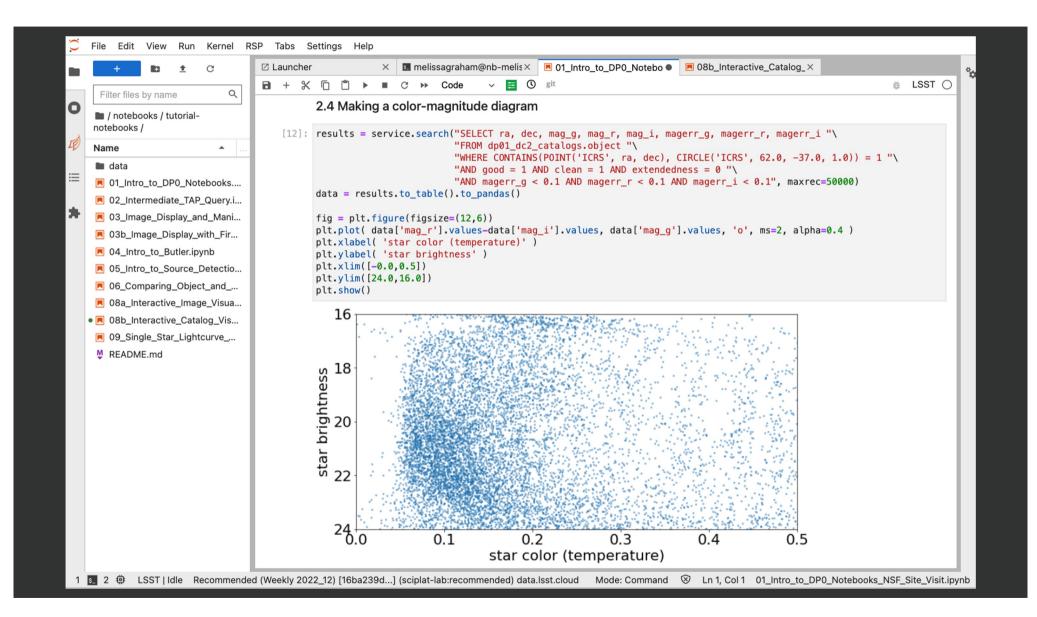
Learn how to programatically access data with Virtual Observatory interfaces

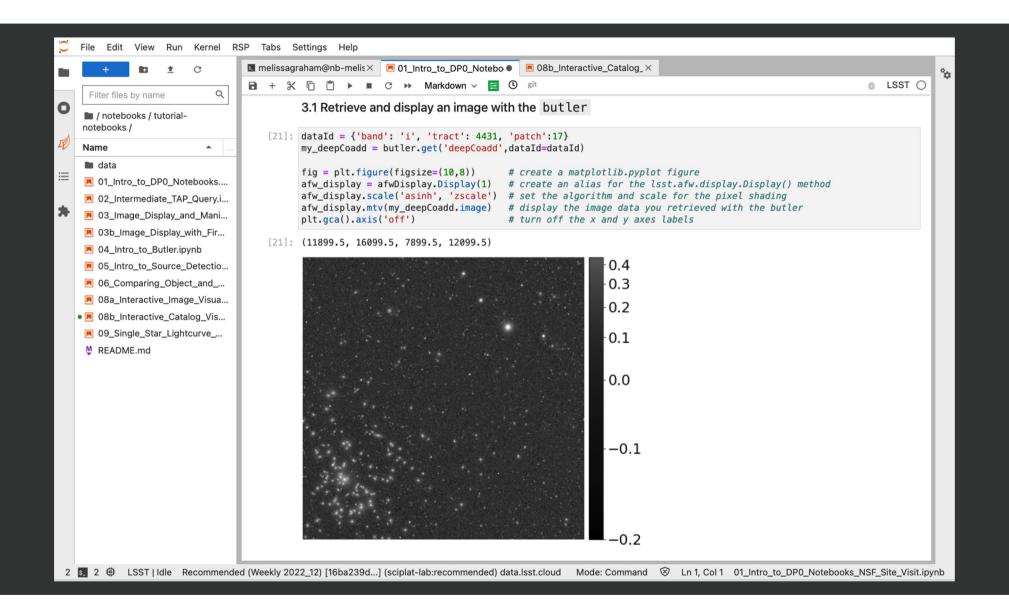
Log in













The butler: a powerful component of the LSST Science Pipelines.



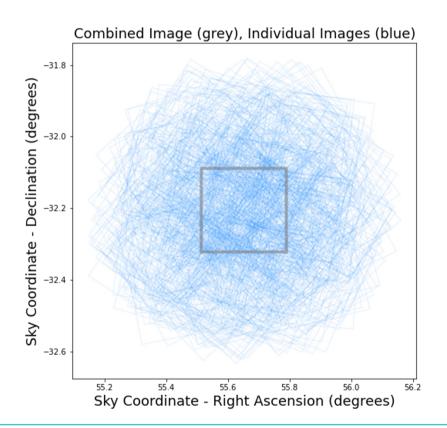
The butler allows users to ignore what is happening "behind the scenes" in terms of where and how the millions of images are stored (or compressed).



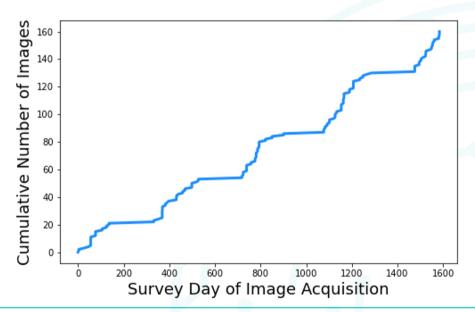
The butler enables users to flexibly query and retrieve *only the data they want* (e.g., full or cropped images, metadata).

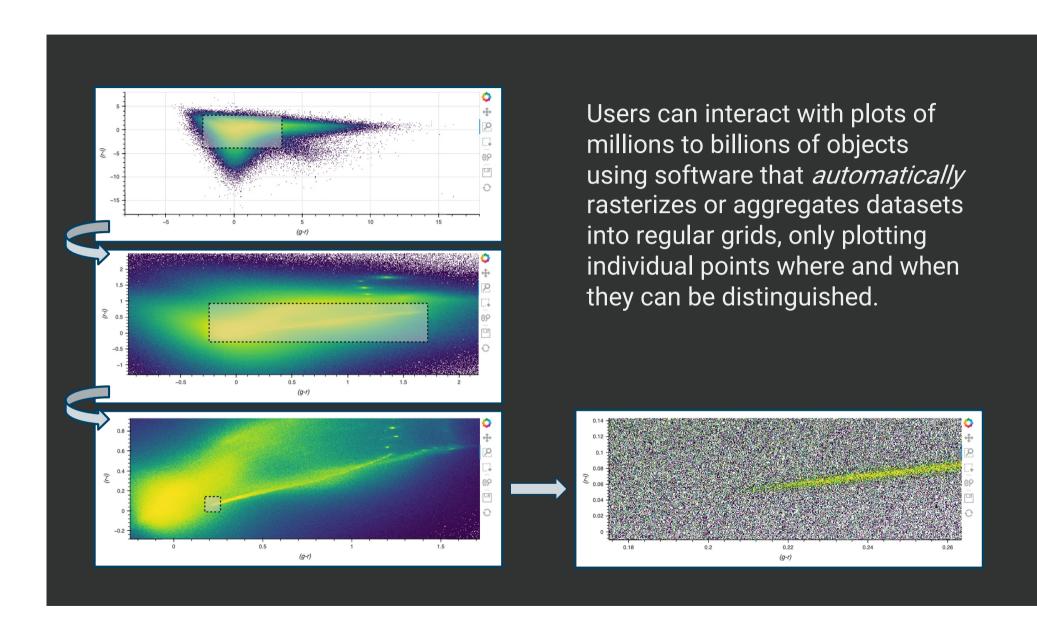


RSP Notebook Aspect



With a couple lines of code and a short wait, users can find all the individual images overlapping any region of sky (left) and metadata like acquisition date (below).





```
[49]: dmap = hv.DynamicMap(update_histogram, streams=[box]).options(
          height=400, width=400)
      datashade(points,
                 cmap=process_cmap("Viridis", provider="bokeh")) * \
      box_plot.options(height=400, width=400) + \
[49]:
                                                                                                     T + 12 10 P 1 □ € 🗘
                               -37
                               -38
                               -39
                            (geb))
                               -42
                               -43
                                        52
                                                54
                                                                58
                                                                                          15
                                                                                                    20
                                                    RA(deg)
      Try changing the box selection and watch as the histogram is recomputed and displayed.
```

⊗ Ln 5, Col 35 08b_Interactive

Mode: Command

LSST | Idle Recommended (Weekly 2022_12) [16ba239d...] (sciplat-lab:recommended) data.lsst.cloud

```
[49]: dmap = hv.DynamicMap(update_histogram, streams=[box]).options(
          height=400, width=400)
      datashade(points,
                cmap=process_cmap("Viridis", provider="bokeh")) * \
      box_plot.options(height=400, width=400) + \
[49]:
                                                                                                     II + 12 10P 1 C 🗘
                              -37
                              -38
                              -39
                            (geb))
                              -42
                              -43
                                       52
                                               54
                                                               58
                                                                                         15
                                                                                                    20
                                                    RA(deg)
      Try changing the box selection and watch as the histogram is recomputed and displayed.
```

Mode: Command 🛞 Ln 5, Col 35 08b_Interactive

LSST | Idle Recommended (Weekly 2022_12) [16ba239d...] (sciplat-lab:recommended) data.lsst.cloud

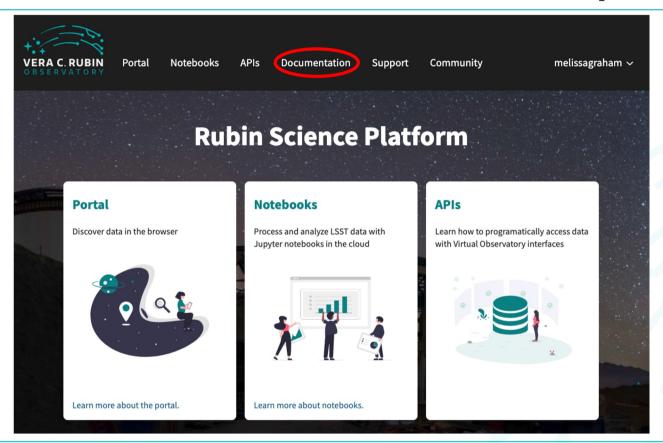
```
[49]: dmap = hv.DynamicMap(update_histogram, streams=[box]).options(
          height=400, width=400)
      datashade(points,
                 cmap=process_cmap("Viridis", provider="bokeh")) * \
      box_plot.options(height=400, width=400) + \
                                                                                                       I + 12 10 P 1 P + 0 + 0
[49]:
                               -37
                               -38
                               -39
                               -42
                               -43
                                        52
                                                 54
                                                                 58
                                                                         60
                                                                                                      20
                                                                                                                 25
                                                     RA(deg)
      Try changing the box selection and watch as the histogram is recomputed and displayed.
   LSST | Idle Recommended (Weekly 2022_12) [16ba239d...] (sciplat-lab:recommended) data.lsst.cloud
                                                                                                                 Mode: Command 🛞 Ln 5, Col 35 08b_Interactive
```



The RSP is an accessible powerful platform built specifically to support big data analysis in the LSST era.









Documentation

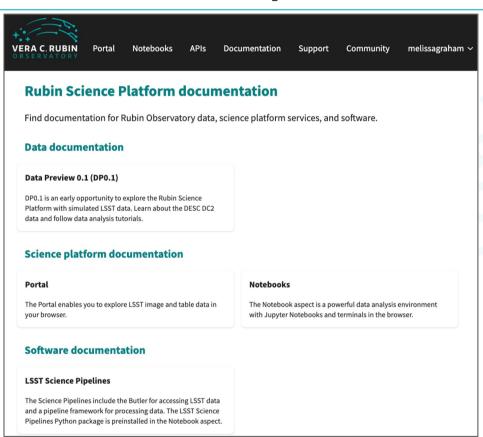
Data Release documentation

- dp0-2.lsst.io
- high-level processing summaries
- data products descriptions & schema
- step-by-step tutorials

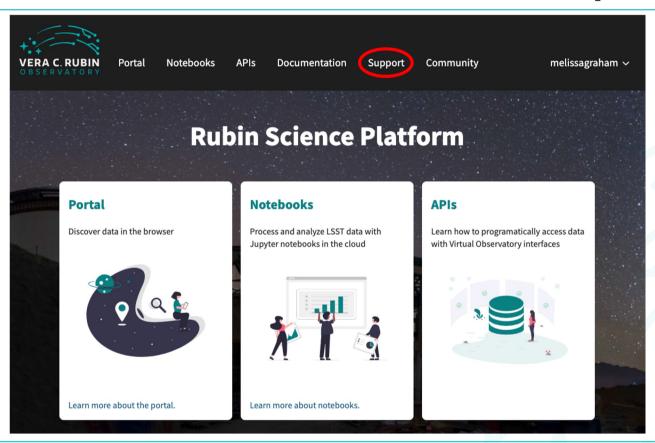
User Guides for the Rubin Science Platform

LSST Science Pipelines documentation

- pipelines.lsst.io
- module descriptions
- use-case examples and FAQs
- software releases & version history



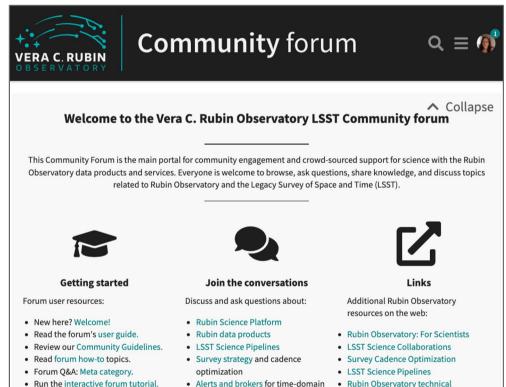






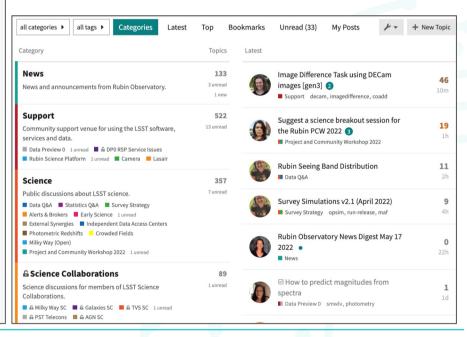
· Review the Terms of Service.

Rubin Science Platform User Experience



astronomy

With *thousands of users*, knowledge bottlenecks are a risk to science. Support must be accessible and timely. A crowd-source model is the only sustainable option.



Vera C. Rubin Observatory | NAM | 11 Jul 2022

documentation