Rubin Observatory Data Preview 1

James Mullaney
The University of Sheffield

VERA C. RUBIN OBSERVATORY

LSST:UK Consortium

NSF NOIR Lab AURA SLAC U.S. DEPARTMENT OF ENERGY UKRI UK Research and Innovation

Who am I, and why am I talking about DP1?

- Lecturer in Extragalactic Astrophysics.
- Since April 2023: LSST:UK's Processing Scientist (50% role). Duties:
  - Quality check data off the telescope and during and after processing;
  - Work with the Data Management team to contribute to the data processing pipelines:
    - Improvements;
    - New Features
    - Bug fixing
  - Was checking data during and after the LSSTComCam observing campaign.
  - Wrote a chunk of the DP1 Data Release paper.

What I will talk about today

- What Data Preview 1 is and isn't.
- What Data Preview 1 contains.
- How to access Data Preview 1.

In-prep DP1 release paper:
https://rtn-095.lsst.io/

What DP1 is…and isn't.

Data Preview 1:

- First on-sky data from the Rubin Observatory to be released for research;
- Was obtained using Rubin's **Commissioning Camera** (ComCam);
  - Same size, weight, detector type, filter complement etc. as the full camara, but containing only **9 detectors, instead of 189**.
  - Used to optically align the Simonyi Telescope and verify it can deliver the required data quality.
- Was obtained during the **48 nights** spanning 24th October - 11th December, 2024.

It is not:

- The data that was shown during the "First Light" event.
- *Exactly* what we expect for full Data Releases.

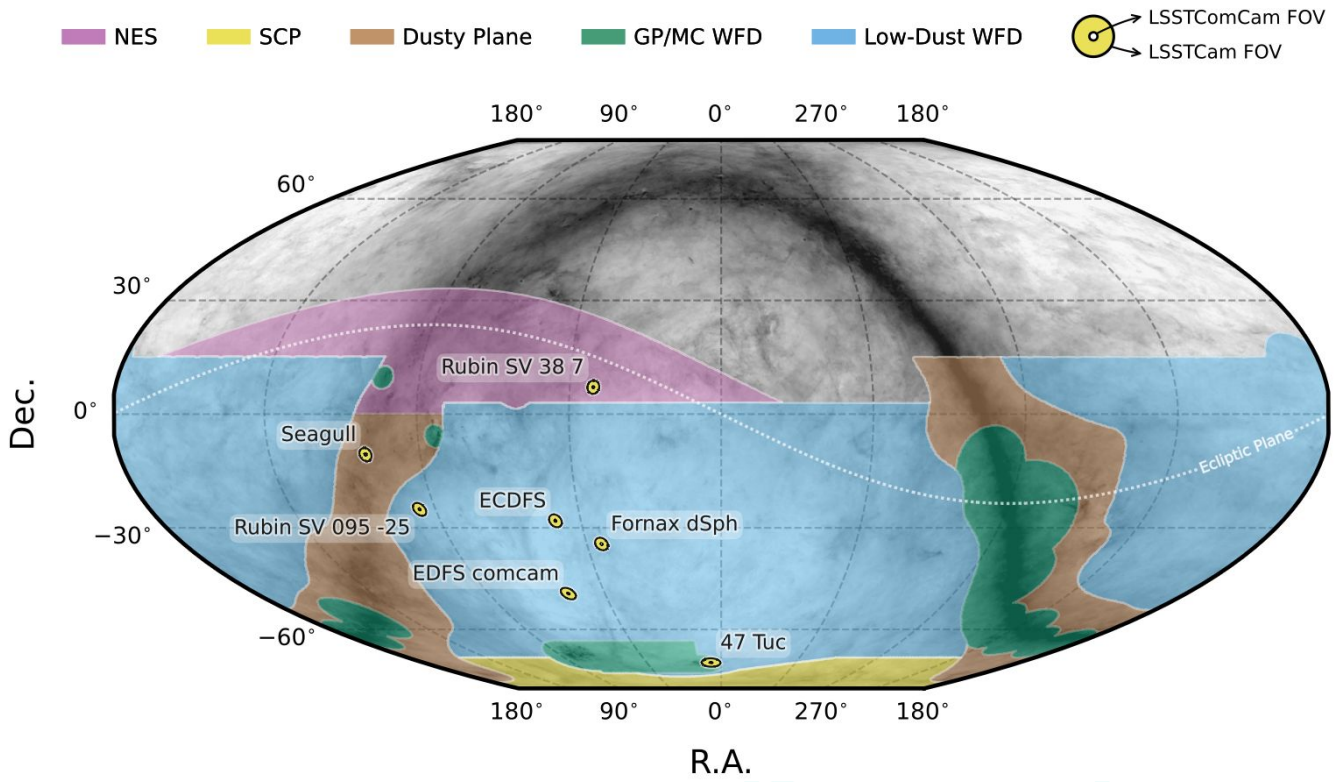
The ComCam Campaign

Targeted 7 fields:

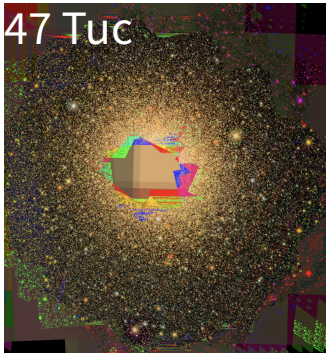
- 47 Tuc
- ECDFS
- EDFS comcam
- Fornax dSph
- Rubin SV 095-25
- Rubin SV 38 7
- Seagull

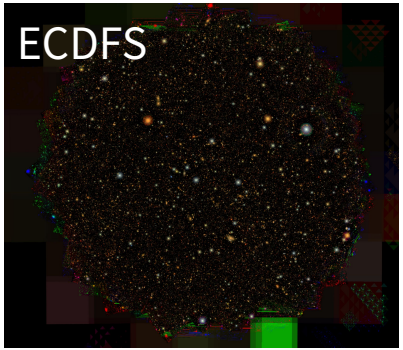
In all six bands:
u, g, r, i, z, y

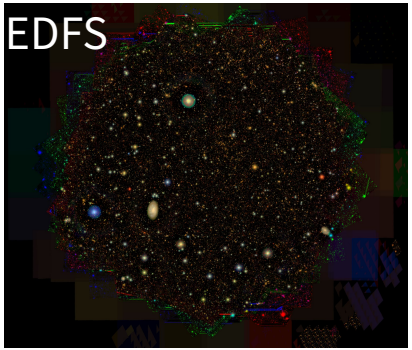
Covered 15 sq. deg.

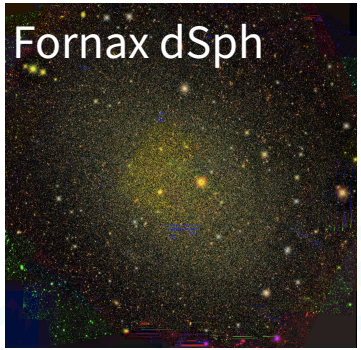


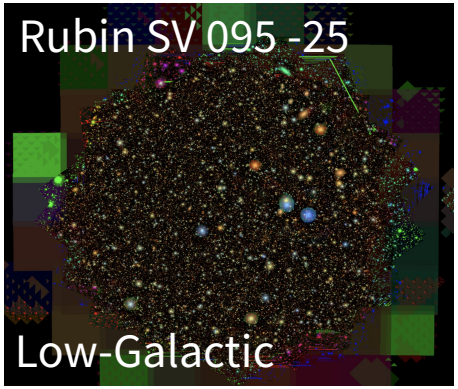
What the fields look like

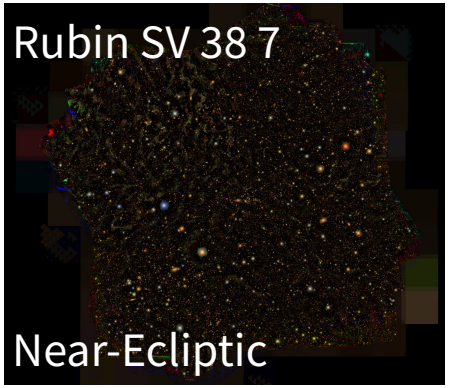


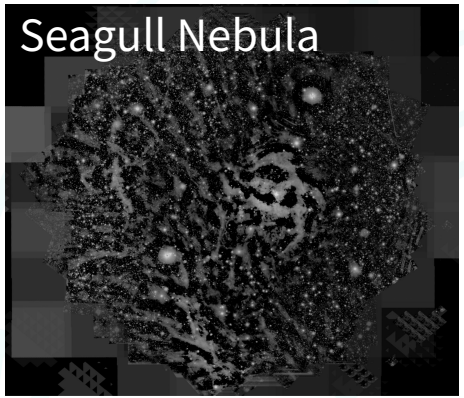












What DP1 contains

DP1 contains data derived from 1792 30s ComCam exposures.

Each exposure resulted in *up to* 9 raw images (and almost all did result in 9).

Number of raw detector images per field and band:

Field Code	Band						Total
	u	g	r	i	z	y	
47_Tuc	54	90	288	171	0	45	648
ECDFS	387	2070	2133	1455	1377	270	7692
EDFS_comcam	180	549	783	378	378	180	2448
Fornax_dSph	0	45	225	108	0	0	378
Rubin_SV_095_-25	297	738	756	207	540	90	2628
Rubin_SV_38_7	0	396	360	495	180	0	1431
Seagull	90	333	387	0	90	0	900
Total	1008	4221	4932	2814	2565	585	16125

What DP1 contains

Data in DP1 fall under five different broad types:

- Images
- Catalogues

Science Data Products

- Maps
- Ancillary Data Products
- Metadata

Aid investigation

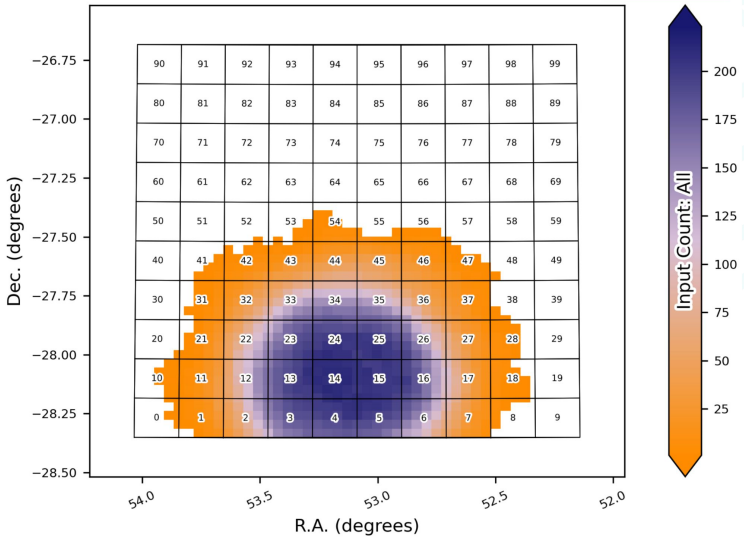
Details of full contents at:
https://dp1.lsst.io/

What DP1 contains - Images

- `raw`: Images directly from the camera, before any processing.
- `visit_image`: Processed & calibrated single epoch images (science, var, mask).
  They are the lowest-level "scientifically useful" images.
- `deep_coadd`: Product of coadding `visit_image`s taken with the same filter.
  Selected to have PSF FWHM < 1.7 arcsec.
- `template_coadd`: Used to make `difference_image`s.
  The third of `visit_image`s with the best seeing are selected.
- `difference_image`: From subtracting `template_coadd`s from `visit_image`s.
  They show what has changed in flux.

Image Health Warnings!

- All processed images contain a mask plane.
  - Provides information about each pixel. Is it associated with a detection, a cosmic ray, a bad pixel….?
  - Make sure you understand and refer to the mask plane if you're working with image data.
- Particularly applies to coadds:
  - Many coadds only have partial coverage.
  - Regions without coverage don't necessarily contain - for example - zeros or NaNs.
  - But they *are* indicated with the `NO_DATA` pixel mask.
  - Ignore the mask planes at your peril!!!



What DP1 contains - Catalogues

- `Source`: Deblended 5-sig detections in `visit_image`s. (46M sources)
- `Object`: Deblended 5-sig detections in `deep_coadd`s. (2.3M objects)
- `ForcedSource`: Forced PSF photometry at object locations (269M entries)
- `DiaSource`: 5-sig detections in `difference_image`s. (3.1M diaSources)
- `DiaObject`: Objects that diaSources are associated with. (1.1M diaObjects)
- `ForcedSourceOnDiaObject`: Forced PSF photometry at diaObject locations (197M)

Solar System:

- `SSSource`: DiaSources that are associated with known Solar System objects (5598)
- `SSObject`: Mapping between Rubin SS Object ID and the IAU designation.

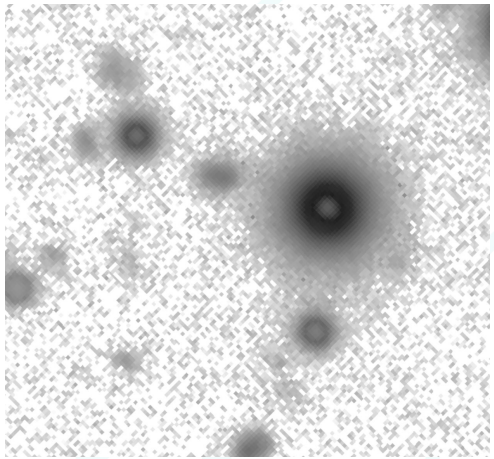
Non-science/not-from-observations:

- `CcdVisit`: Information about the exposures. Pointing, PSF, etc. (16071 entries)
- `Calibration`: The catalogue used for astrom. and photom. calibration.

Catalogue Health Warnings!

Flags:

- Catalogues also contain flags.
- Indicate whether a detection/measurement is associated with a bad pixel, poorly-defined PSF etc.
- Make sure you understand and exclude appropriately flagged sources. e.g., `~base_psfFlux_Flag`.

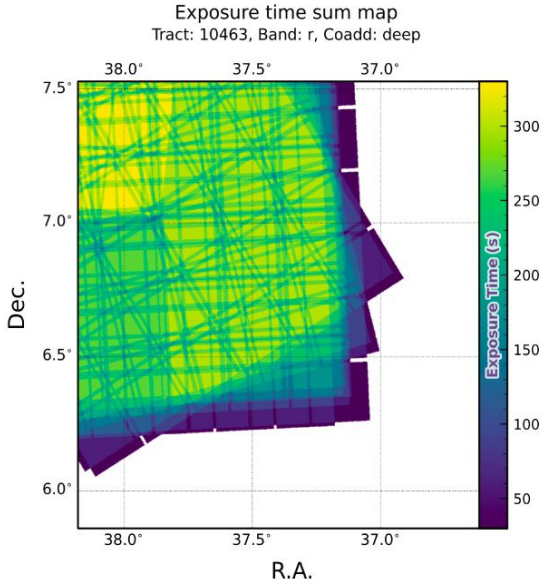


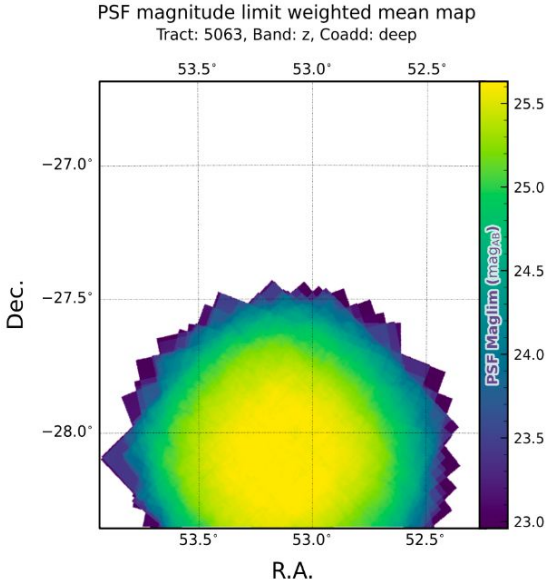
Lightcurves:

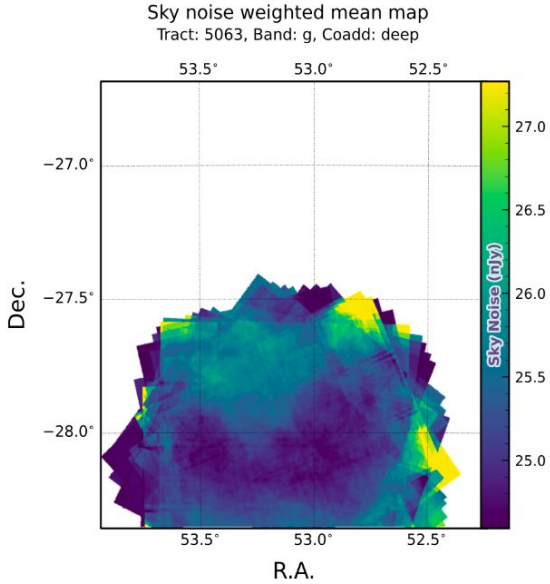
For lightcurves, we strongly recommend using the `psfDiffFlux` column in `ForcedSource`.

Maps

Maps are non-science-level images. Examples include:







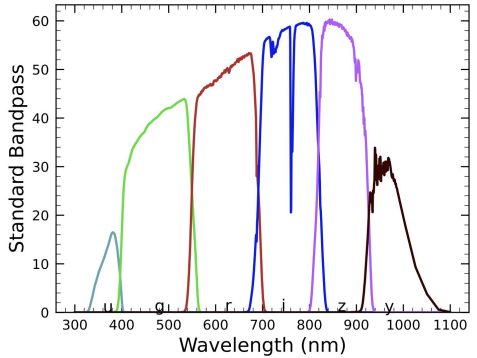
Metadata & Ancillary Data Products

Metadata: `visit` table contains data for each exposure.

Differs from `ccdVisit`: per exposure, as opposed to per exposure-detector.

Ancillary Data Products:

- *Config files*: So users can know how the pipeline was configured.
- *log and metadata files*: Produced during processing, allows users to check why a certain processing failed, for example.
- *calibration data products*: Bias, flats, darks, brighter-fatter kernels, charge transfer inefficiencies…
- *whole system bandpass throughputs*…



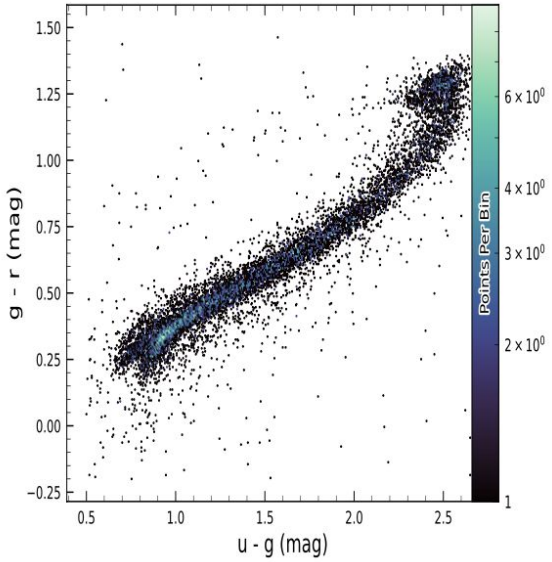
DP1 Data Quality - 5sigma depths

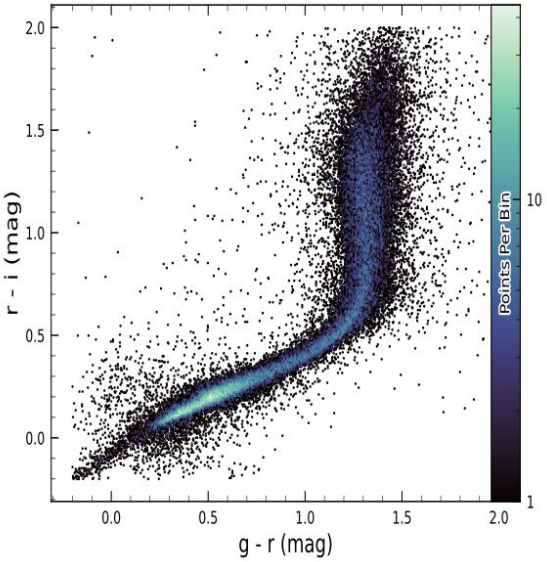
Table 1. Median 5σ coadd detection limits per field and band.

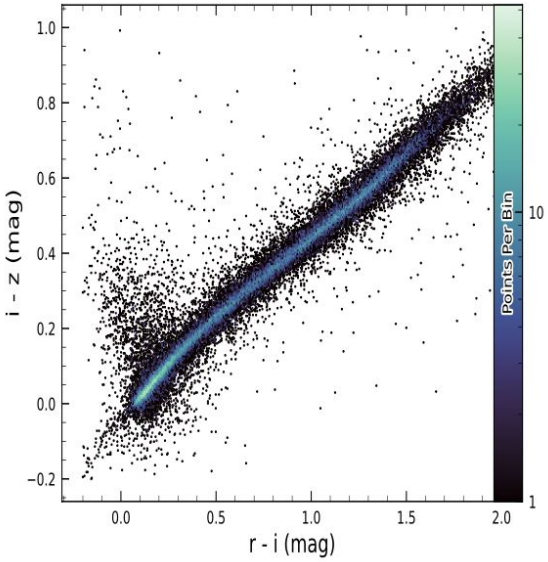
Field Code	Band					
	u	g	r	i	z	y
47_Tuc	-	24.03	24.24	23.90	-	21.79
ECDFS	24.55	26.18	25.96	25.71	25.07	23.10
EDFS_comcam	23.42	25.77	25.72	25.17	24.47	23.14
Fornax_dSph	-	24.53	25.07	24.64	-	-
Rubin_SV_095_-25	24.29	25.46	24.95	24.86	24.32	22.68
Rubin_SV_38_7	-	25.46	25.15	24.86	23.52	-
Seagull	23.51	24.72	24.19	-	23.30	-
10-yr depth	26.1	27.4	27.5	26.8	26.1	24.9

Unofficial!

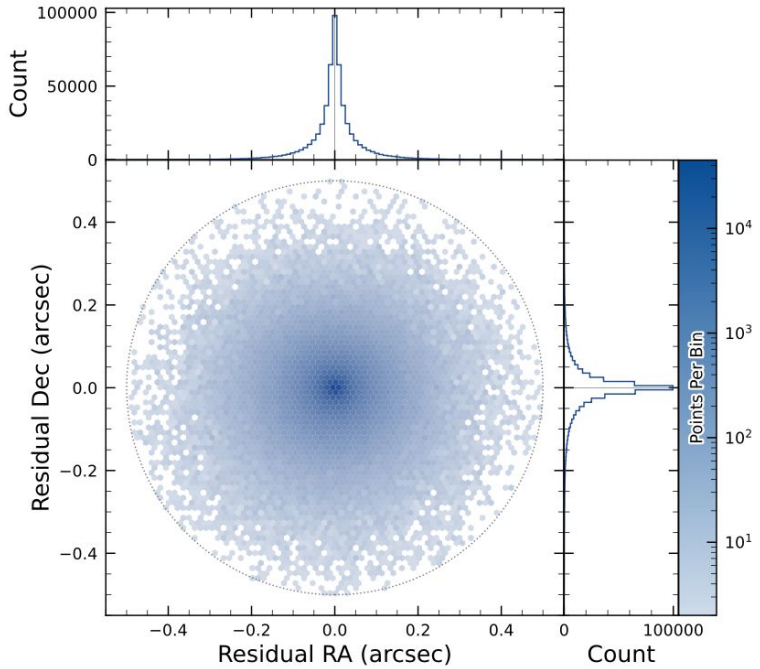
Photometry checks



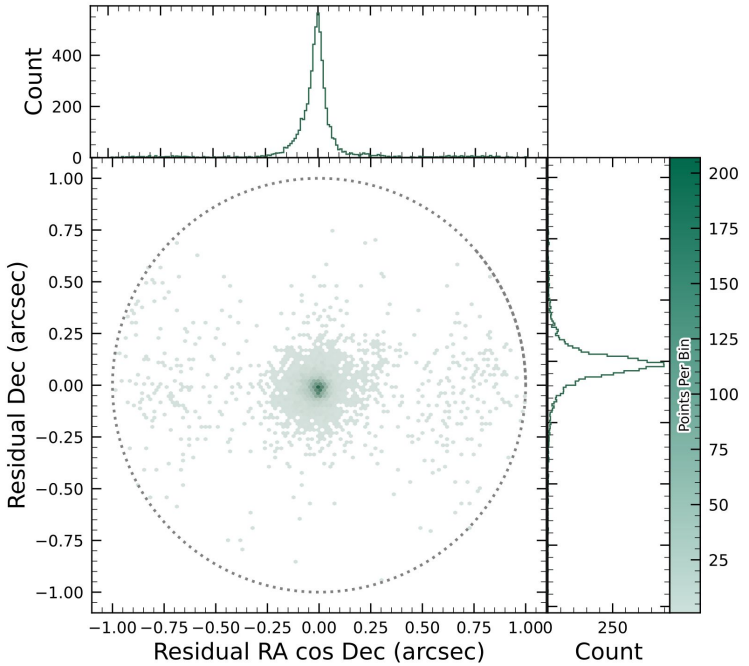




Astrometry checks

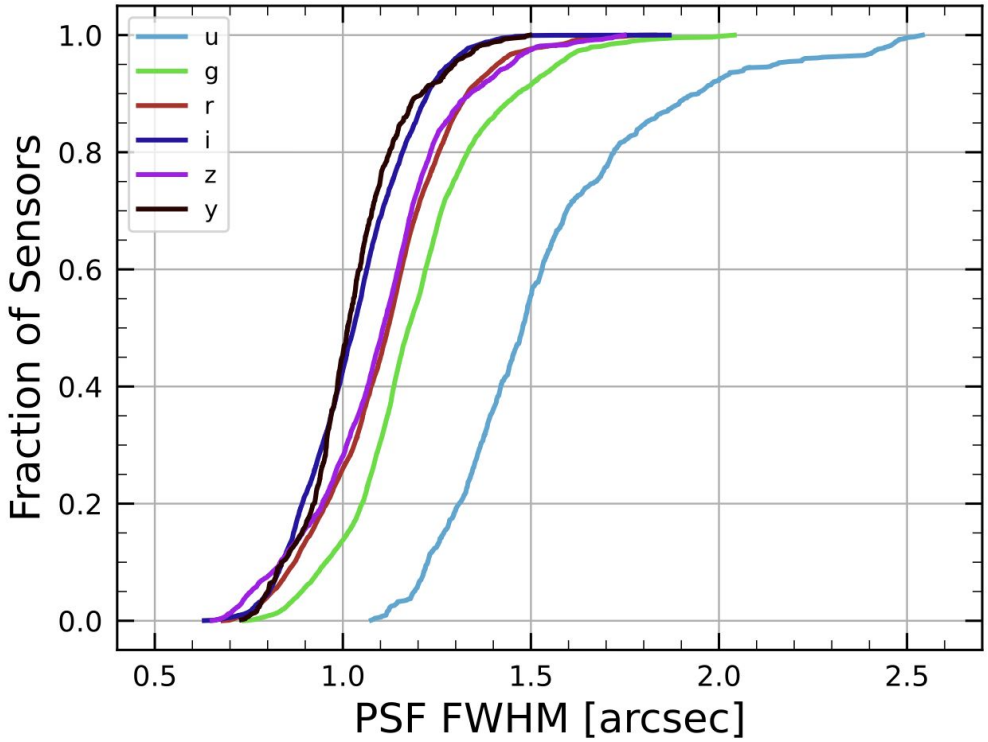


ECDFS astrometric residuals



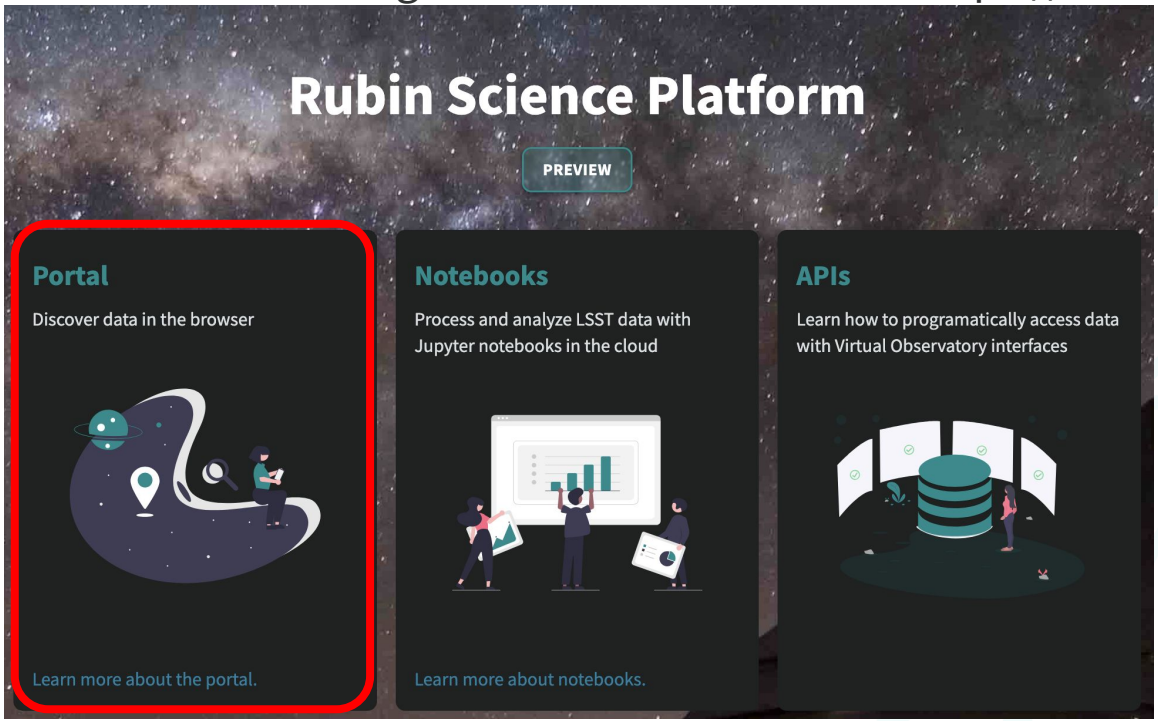
SSObject astrometric residuals

Point Spread Function

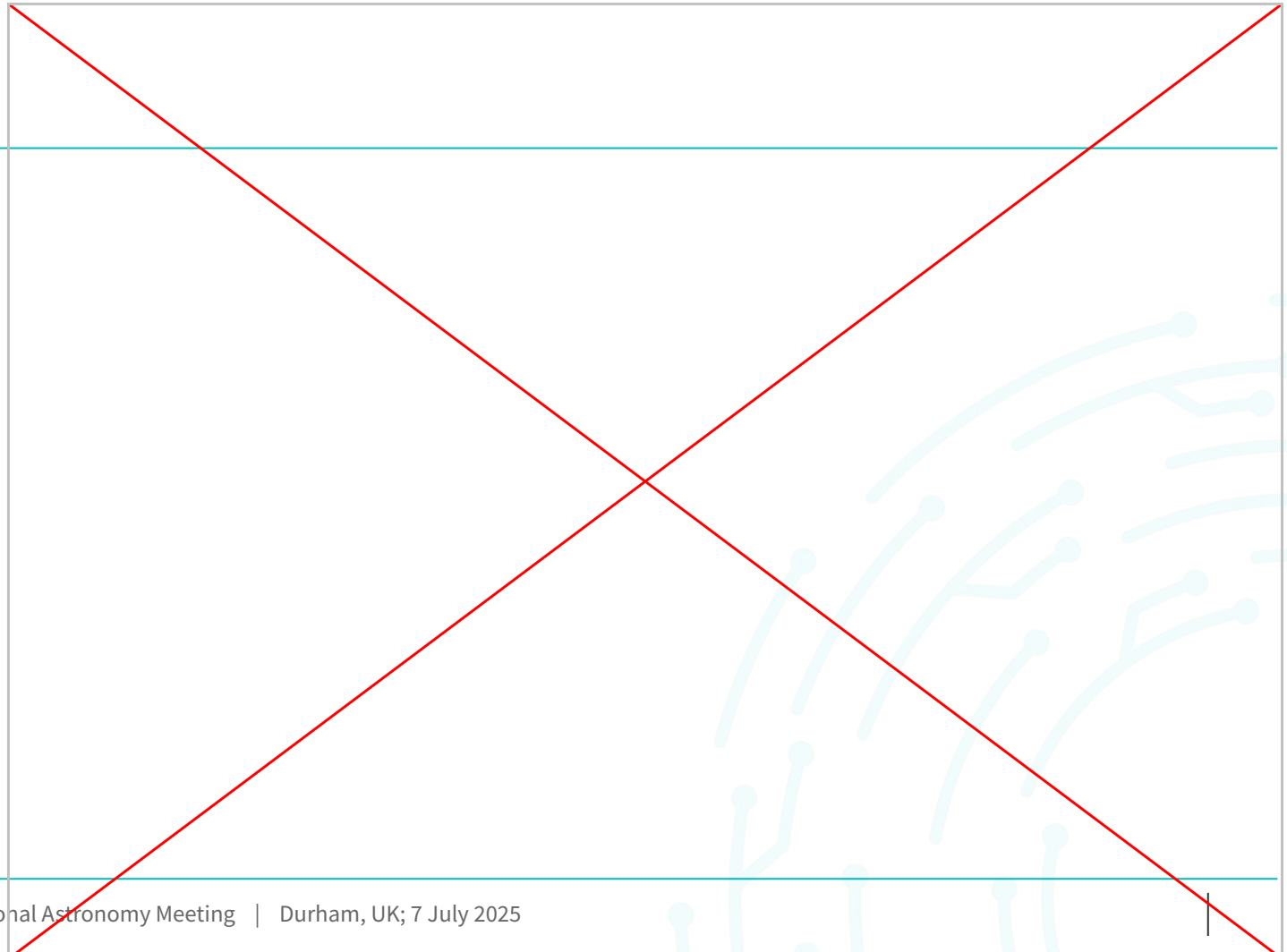


Accessing the data

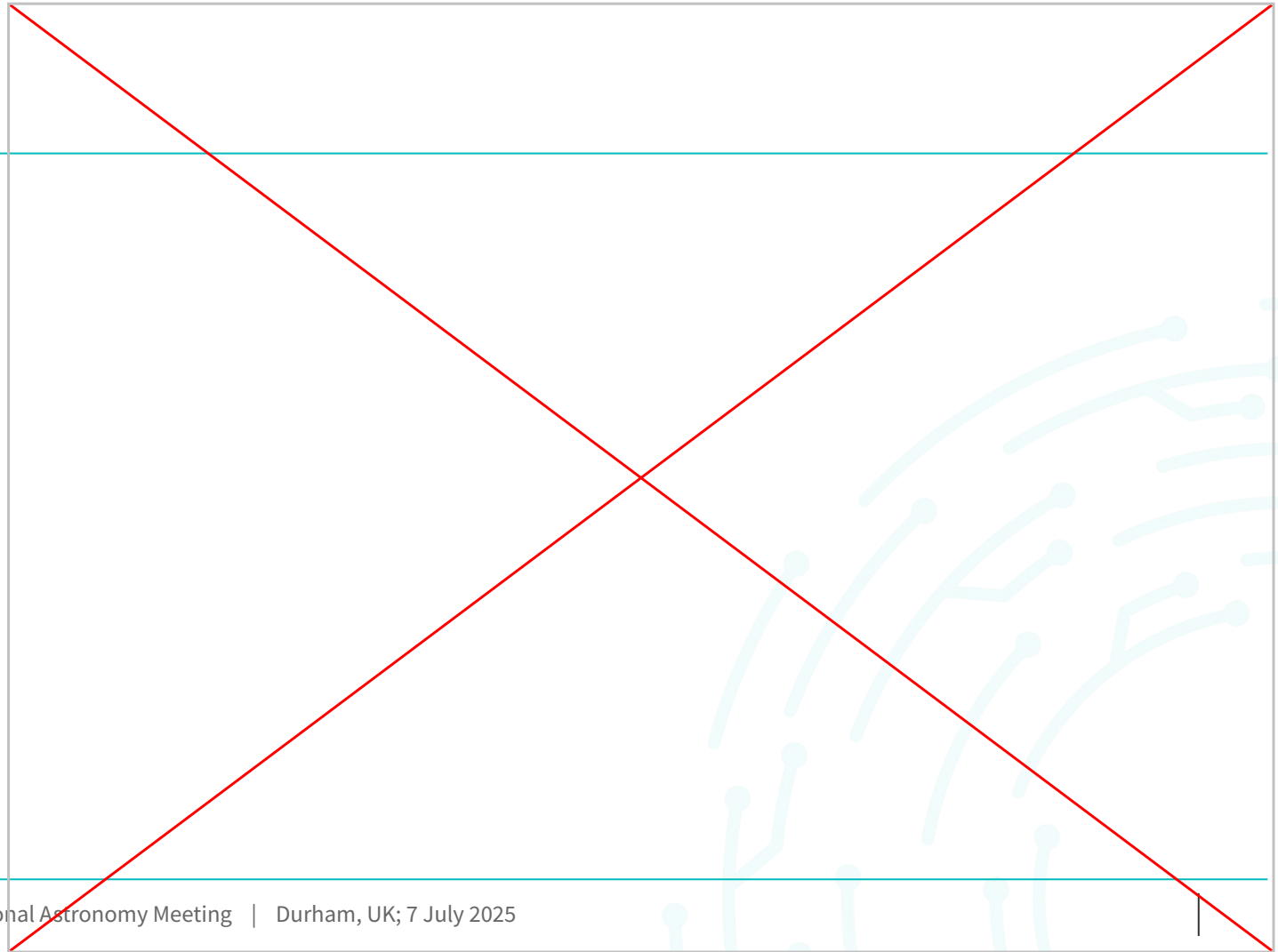
DP1 data is available **now** to Data Rights holders. Access is via https://data.lsst.cloud:



Video of accessing HiPS maps



Video of accessing Catalogues in Portal



Accessing the data

DP1 data is available **now** to Data Rights holders. Access is via https://data.lsst.cloud:

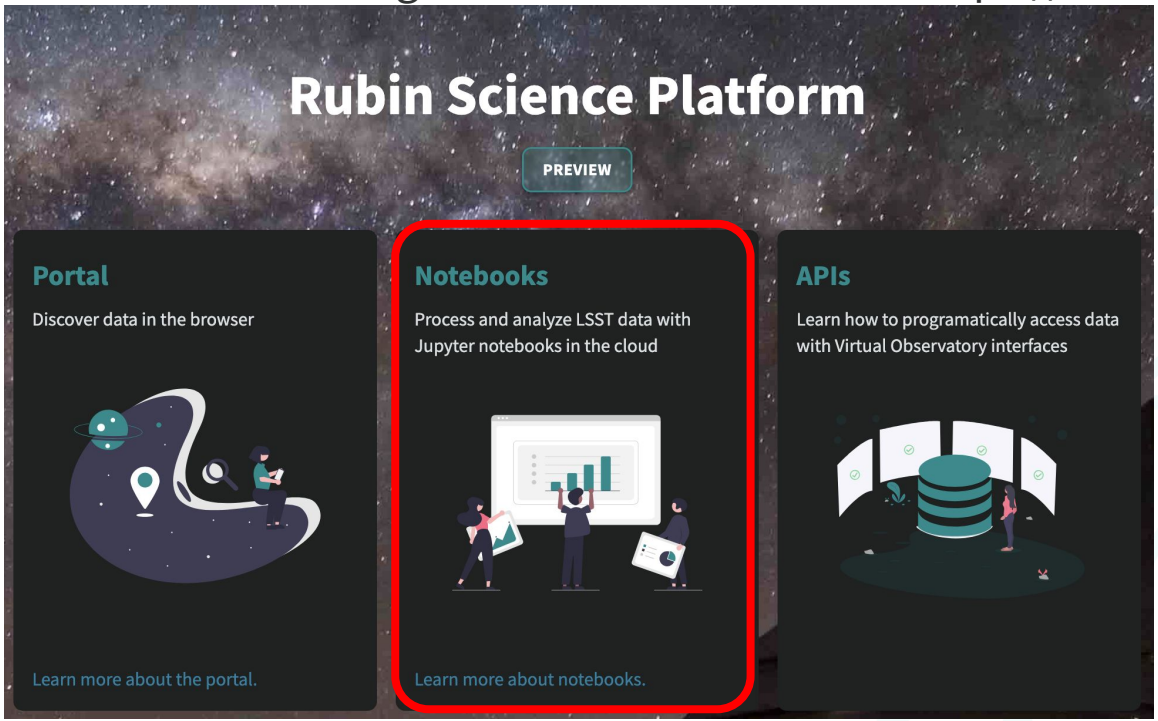
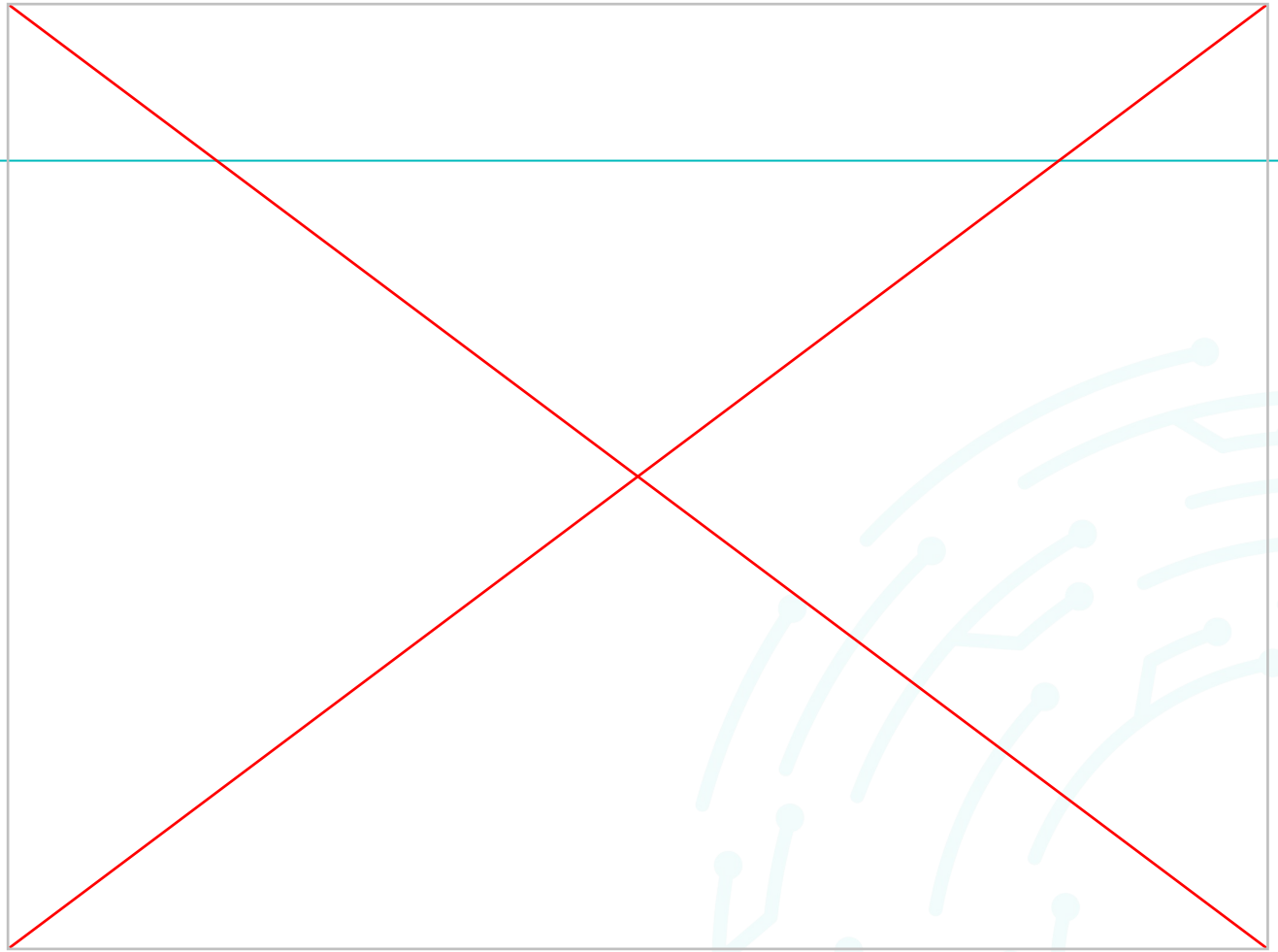


Table Access Platform (TAP)

- Query the same tables as the Portal Aspect, but in a notebook setting.
- It's the recommended way of programmatically accessing table data.
- ADQL - SQL with astronomy features - queries *all* objects, source etc. at once.
- Once searched-for and retrieved, data can be manipulated and analysed.
- It is faster - but less feature-rich and flexible - than using the Data Butler.

Video of accessing catalogues via TAP in Notebook.



The Data Butler

- Use the Data Butler to retrieve pixel-level data for analysis.
- But it can retrieve table data, too.
- It's very powerful and feature-rich.
- Is more flexible than TAP, but less efficient for querying table data.
- Can be used to acquire broader types of information, such as tract and patch information.

Video of accessing a coadd via the Data Butler

Code | Notebook | LSST

Using the Data Butler to retrieve an image

Import necessary modules and set up plotting

```
[ ]: import lsst.daf.butler as dafButler
     import matplotlib.pyplot as plt
     import lsst.afw.display as afwDisplay

     afwDisplay.setDefaultBackend('matplotlib')
     plt.rcParams['figure.figsize'] = (8.0, 8.0)

     def plotImage(image, favPx=None):
         fig = plt.figure()
         display = afwDisplay.Display(frame=fig)
         display.scale('asinh','zscale')
         display.mtv(image)
         if favPx is not None:
             display.dot('o', favPx[0], favPx[1], size=100, ctype='orange')
```

Set up the Data Butler

```
[ ]: butler = dafButler.Butler('dp1', collections=['LSSTComCam/DP1'])
     registry = butler.registry
```

Retrieve the coadd image data...

```
[ ]: id =  {'tract': 5063, 'patch':55, 'band':'g', 'skymap':'lsst_cells_v1'}
     refs = list(registry.queryDatasets('deep_coadd', dataId=id))
     coadd = butler.get(refs[0])
```

Summary

- Data Preview 1 was made available to Data Rights holders on the 30th Nov 2025.
- DP1 is the first science-grade data to be released by the Rubin Observatory.
- It contains data from the ComCam campaign: 48 nights, 7 fields, 15 sq. deg.
- Excellent data quality in terms of photometry, astrometry, PSF, colours.
    - But we can't check *everything* - so we want to hear from the community.
- There are multiple ways to explore the data:
    - Portal Aspect to visualise images and catalogs;
    - Notebook Aspect to manipulate and analyse data:
        - TAP for straightforward catalogue searches;
        - Butler for image and more complex catalogue searches.
- Go Forth and Discover Stuff!!