LAPP
Laboratoire d'Annecy de Physique des Particules

- Dominique BOUTIGNY
- Sabine ELLES
- Marie PATUREL

ESAP Training Workshop

Rubin Science Platform

ESCAPE meeting - November 22, 2022

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

Credit

The Rubin Science Platform (RSP) is developed by the SQuaRE* team, Rubin Data Management funded by NFS and DOE

All the RSP code is Open Source and freely reusable: https://github.com/lsst-sqre

(*) SQuaRE = **S**cience **Q**uality and **R**eliability **E**ngineering team

Thanks to Gabriele Mainetti who is deploying the RSP at CC-IN2P3

Funded by the European Union's Horizon 2020 - Grant N° 824064

Rubin Observatory

- 8.4 m optical telescope
- 3.2 Gpixel camera
- 9.6 deg² field of view
- 6 pass band filters (u,g,r,i,z,y)

Global survey of the sky 2023-2032

- 20 000 deg²
- 1 image every 40s every night for 10 years
- Full sky surveyed every 3-4 nights
- Deep Drilling Field visited several times every nights

Funded by the European Union's Horizon 2020 - Grant N° 824064

Rubin Data Management

Borrowed to Leanne Guy (Rubin DM Scientist)

Data Release Data Products
via Annual Data Releases

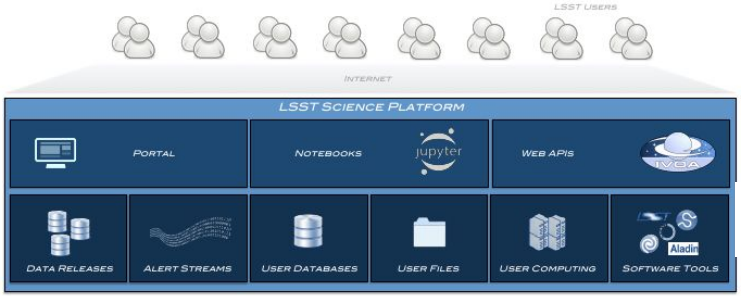
11 Data Releases in 10 years
Final database catalog: 15 PB

20TB raw data/night
(with calibration exposures)

Prompt Data Products
via nightly Alert Streams

Average ~ 10 million/night
Real-time latency: 60sec

Rubin Science Platform

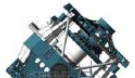


Data Releases
(current & previous)

Alerts database

Alert streams → Brokers

Data access via Data Access Centres & Services



LSST Key Numbers

Rubin Data Management

Borrowed to Leanne Guy (Rubin DM Scientist)

Data Release Dat
via Annual D

..a Releases in 10 years
..al database catalog: 15 PB

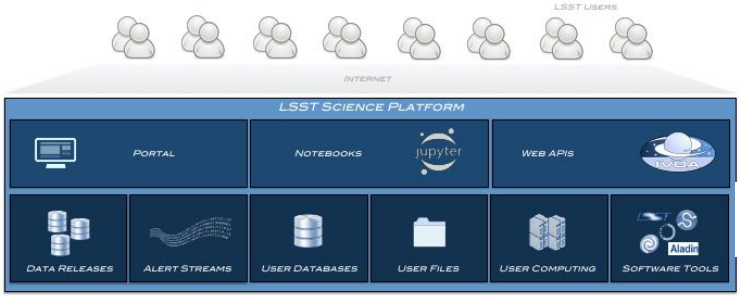
France / IN2P3 is responsible of 50% of the Data Release Processing (DRP) → CC-IN2P3 (Lyon)

20TB raw data/night
(with calibration exposures)

Prompt Data Products
via nightly Alert Streams

Average ~ 10 million/night
Real-time latency: 60sec

Rubin Science Platform

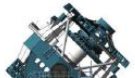


Data Releases
(current & previous)

Alerts database

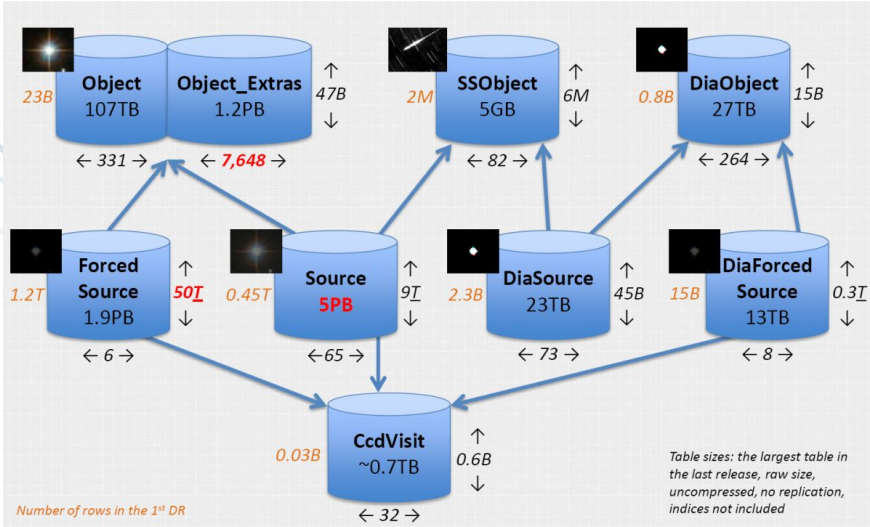
Alert streams → Brokers

Data access via Data Access Centres & Services



The Rubin Dataset

- 2.75 10^6 visits (2 snaps - 189 CCD)
- 60 PB raw data / year
- Final dataset : 500 PB



- Peak computing power : 1.8 PFlops

Qserv massively parallel shared nothing distributed SQL database

- 37 10^9 sources
- 3. 10^9 well measured galaxies (position, shape, colors)
- Final catalog (Data Release 11) in database : 15 PB
  - catalogs also available in *parquet* format

Funded by the European Union's Horizon 2020 - Grant N° 824064

Some Rubin Specificities

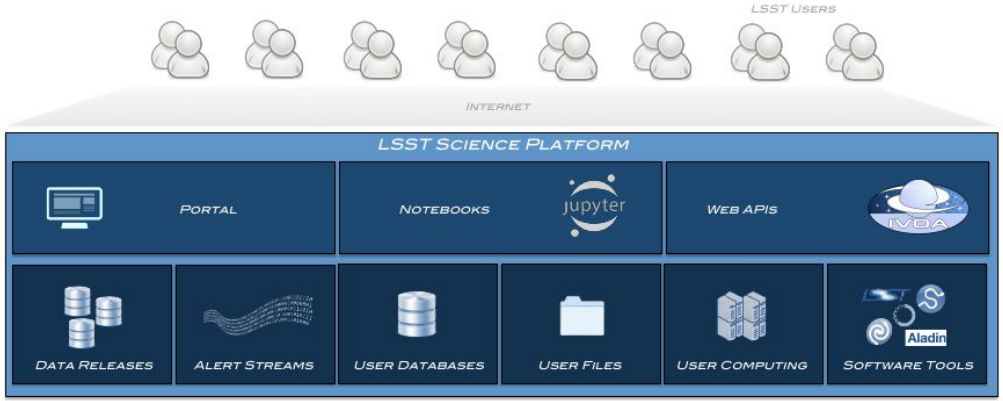
- Rubin will have to serve a large community of users organized in several Science Collaborations (Solar System, Milky Way, Galaxies, Dark Energy, Transients,...)
- Some collaborations (i.e. Dark Energy Science Collaboration) will have to handle extra processing using specific algorithms

  ⇒ Produce extra datasets

- Rubin data will be proprietary for 2 years after Data Release
  - Alerts are immediately public (through brokers)
- LSST code is fully open source (see: https://github.com/lsst) and re-usable
  - *Subaru / HSC and Rubin are sharing the same Data Release Processing code*

Funded by the European Union's Horizon 2020 - Grant N° 824064

Rubin Science Platform (RSP)



- RSP(s) are being deployed at International Rubin Data Access Centers
- 3 main user interfaces:
  - Web portal - Jupyter notebooks - Web APIs
- Provide access to Rubin Data Releases (including images) and alerts streams

Funded by the European Union's Horizon 2020 - Grant N° 824064

Relation with ESCAPE

- The RSP is not limited to the Rubin Observatory
- Can access any dataset / catalog through IVOA
  - Query through TAP service
- Fully open source
- Modular design allows to extend it and to add / replace modules

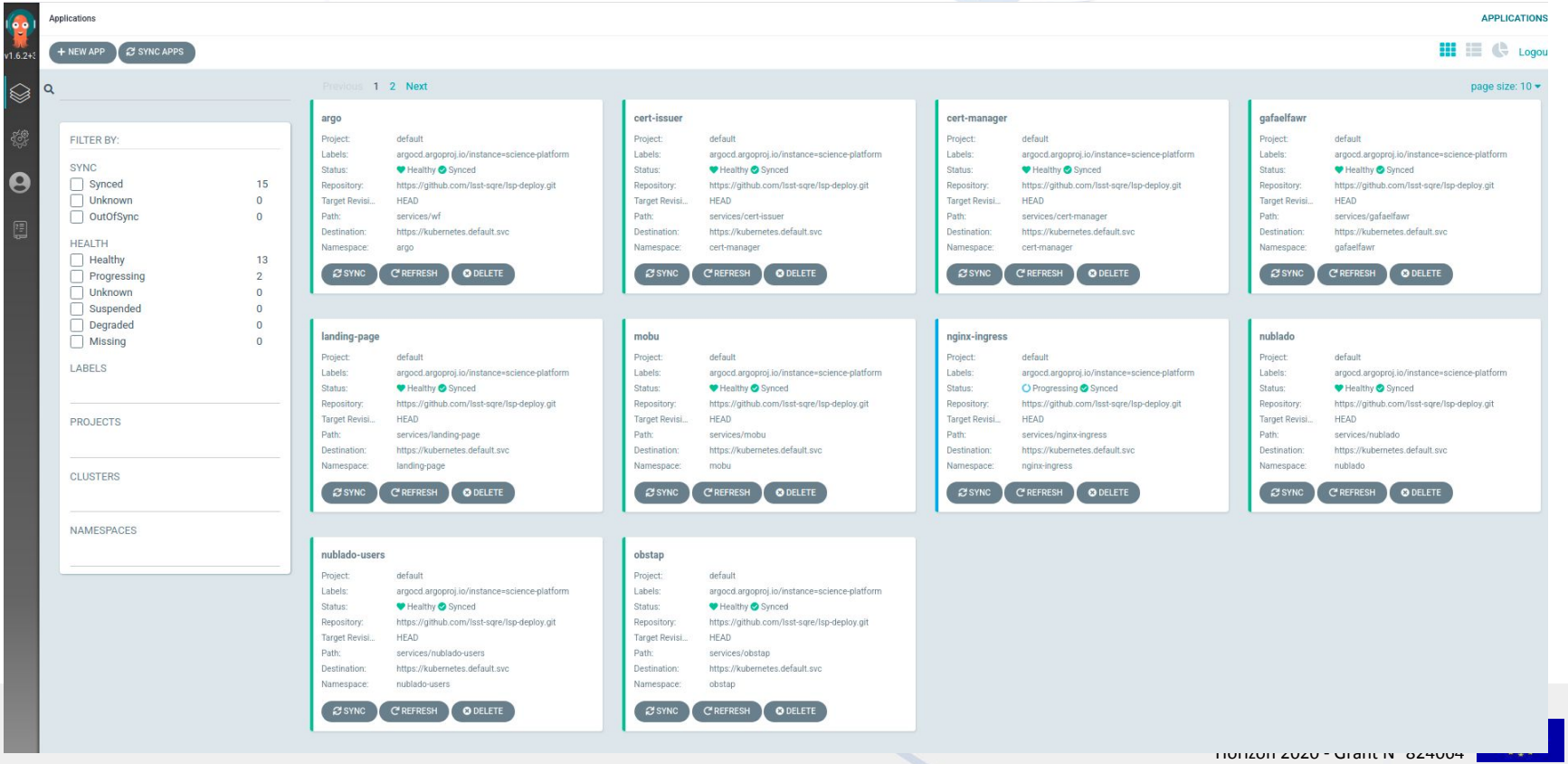
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Main characteristics

- Based on Kubernetes + HELM + argo
  - Container management system
  - Allows to create a virtual complex cluster (including network) with several interacting components
- k8s secrets kept in Vault
- github, openid and ldap authentication are currently supported
- **Infrastructure as Code**
  - The k8s infrastructure is described in configuration files kept on github
    - 1 set of configuration files per site
  - A change in configuration (commit / push) is automatically detected by the platform
    - trigger resynchronization of the impacted k8s components

Funded by the European Union's Horizon 2020 - Grant N° 824064

k8s services managed through argo CD → dash board



Horizon 2020 - Grant N° 824064

Deployment status

- Official central RSP available on Google cloud
  - Already opened to ~600 Rubin "delegates" for testing

- Deployments at LAPP (in the WP5 framework) and at CC-IN2P3 (in the Rubin framework)

- Complex to deploy, especially on real hardware (“on premise”)
  - Kubernetes has been originally designed for cloud environments
  - Several services have to be deployed manually while they just show up automatically on commercial clouds (Google)
- But very robust once deployed
  - Scalability (designed to scale up to 10 000 users)
  - Security

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

Deployment status

- Official central RSP available on Google cloud
  - Already opened to ~600 Rubin delegates for testing
- Deployments at LAPP (in the WP5 framework) and CC-IN2P3 (in the Rubin framework)
- Complex to deploy, especially on hardware ("on premise")
  - Kubernetes has been originally designed for cloud environments
  - Several services have to be deployed manually while they just show up automatically on commercial clouds (Google)
- But very robust once deployed
  - Scalability (designed to scale up to 10 000 users)
  - Security

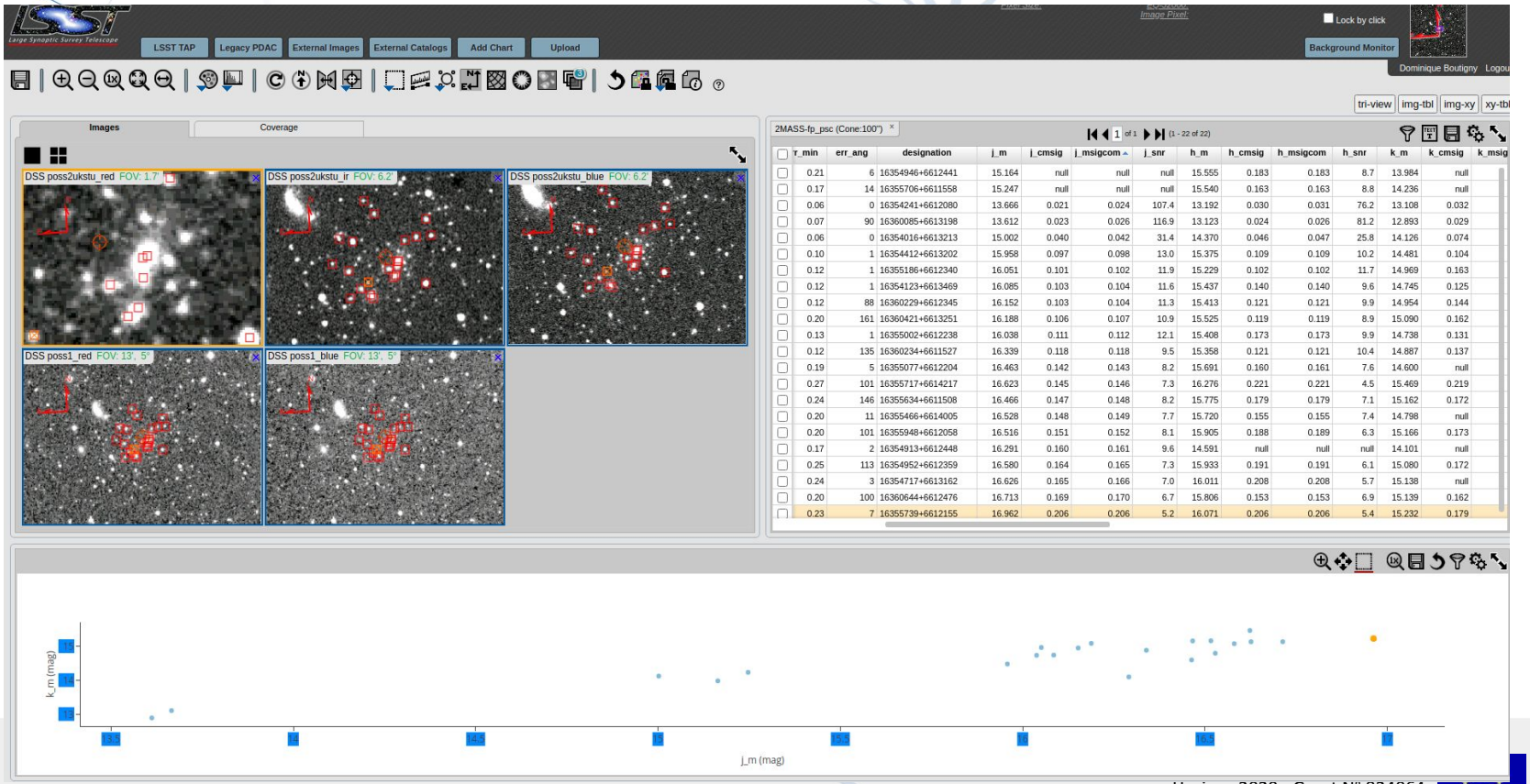
DevOps expertise is becoming crucial !

Funded by the European Union's Horizon 2020 - Grant N° 824064

TAP / Firefly

Firefly is an interactive Java based visualization tool developed at Caltech-IPAC

https://github.com/Caltech-IPAC/firefly



r_min	err_ang	designation	j_m	j_cmsig	j_msigcom	j_snr	h_m	h_cmsig	h_msigcom	h_snr	k_m	k_cmsig	k_msig
0.21	6	16354946+6612441	15.164	null	null	null	15.555	0.183	0.183	8.7	13.984	null	
0.17	14	16355706+6611558	15.247	null	null	null	15.540	0.163	0.163	8.8	14.236	null	
0.06	0	16354241+6612080	13.666	0.021	0.024	107.4	13.192	0.030	0.031	76.2	13.108	0.032	
0.07	90	16360085+6613198	13.612	0.023	0.026	116.9	13.123	0.024	0.026	81.2	12.893	0.029	
0.06	0	16354016+6613213	15.002	0.040	0.042	31.4	14.370	0.046	0.047	25.8	14.126	0.074	
0.10	1	16354412+6613202	15.958	0.097	0.098	13.0	15.375	0.109	0.109	10.2	14.481	0.104	
0.12	1	16355186+6612340	16.051	0.101	0.102	11.9	15.229	0.102	0.102	11.7	14.969	0.163	
0.12	1	16354123+6613469	16.085	0.103	0.104	11.6	15.437	0.140	0.140	9.6	14.745	0.125	
0.12	88	16360229+6612345	16.152	0.103	0.104	11.3	15.413	0.121	0.121	9.9	14.954	0.144	
0.20	161	16360421+6613251	16.188	0.106	0.107	10.9	15.525	0.119	0.119	8.9	15.090	0.162	
0.13	1	16355002+6612238	16.038	0.111	0.112	12.1	15.408	0.173	0.173	9.9	14.738	0.131	
0.12	135	16360234+6611527	16.339	0.118	0.118	9.5	15.358	0.121	0.121	10.4	14.887	0.137	
0.19	5	16355077+6612204	16.463	0.142	0.143	8.2	15.691	0.160	0.161	7.6	14.600	null	
0.27	101	16355717+6614217	16.623	0.145	0.146	7.3	16.276	0.221	0.221	4.5	15.469	0.219	
0.24	146	16355634+6611508	16.466	0.147	0.148	8.2	15.775	0.179	0.179	7.1	15.162	0.172	
0.20	11	16355466+6614005	16.528	0.148	0.149	7.7	15.720	0.155	0.155	7.4	14.798	null	
0.20	101	16355948+6612058	16.516	0.151	0.152	8.1	15.905	0.188	0.189	6.3	15.166	0.173	
0.17	2	16354913+6612448	16.291	0.160	0.161	9.6	14.591	null	null	null	14.101	null	
0.25	113	16354952+6612359	16.580	0.164	0.165	7.3	15.933	0.191	0.191	6.1	15.080	0.172	
0.24	3	16354717+6613162	16.626	0.165	0.166	7.0	16.011	0.208	0.208	5.7	15.138	null	
0.20	100	16360644+6612476	16.713	0.169	0.170	6.7	15.806	0.153	0.153	6.9	15.139	0.162	
0.23	7	16355739+6612155	16.962	0.206	0.206	5.2	16.071	0.206	0.206	5.4	15.232	0.179	

Horizon 2020 - Grant N° 824064

nublado2 Jupyter platform

File Edit View Run Kernel Rubin Tabs Settings Help

test.ipynb | cluster.ipynb | DC2_catalog_access.ipy | Firefly.ipynb | boutigny@nb-boutigny:~ | 07a_DiaObject_Samples

Markdown | git | LSST

Filter files by name

/ tutorial-notebooks /

Name	Last Modified
data	13 minutes ago
01_Introduction_to_DP02.ipynb	13 minutes ago
02_Catalog_Queries_with_TAP.ipynb	13 minutes ago
03a_Image_Display_and_Manipulation…	13 minutes ago
03b_Image_Display_with_Firefly.ipynb	13 minutes ago
03c_Survey_Property_Maps.ipynb	13 minutes ago
04a_Introduction_to_the_Butler.ipynb	13 minutes ago
04b_Intermediate_Butler_Queries.ipynb	13 minutes ago
05_Introduction_to_Source_Detection.i…	13 minutes ago
06a_Interactive_Image_Visualization.ip…	13 minutes ago
06b_Interactive_Catalog_Visualization.…	13 minutes ago
07a_DiaObject_Samples.ipynb	10 minutes ago
07b_Variable_Star_Lightcurves.ipynb	13 minutes ago
08_Truth_Tables.ipynb	13 minutes ago
09a_Custom_Coadd.ipynb	13 minutes ago
09b_Custom_Coadd_Sources.ipynb	13 minutes ago
README.md	13 minutes ago

```
                          "rPSFluxMAD, rPSFluxChi2, rPSFluxNdata, rPSFluxSkew, "
                          "rPSFluxStetsonJ, rPSFluxPercentile05, rPSFluxPercentile25, "
                          "rPSFluxPercentile50, rPSFluxPercentile75, rPSFluxPercentile95, "
                          "rTOTFluxMean, rTOTFluxSigma "
                          "FROM dp02_dc2_catalogs.DiaObject ",
                          maxrec=100000)

DiaObjs = results.to_table()
del results
```

```
CPU times: user 5.31 s, sys: 192 ms, total: 5.5 s
Wall time: 1min 1s
```

Plot the number of DiaSources per DiaObject

Plot the distribution of the number of DiaSources per DiaObject (i.e., the total number of difference-image detections in any filter; at left), and the distribution of the number of r-band DiaSources per DiaObject (at right).

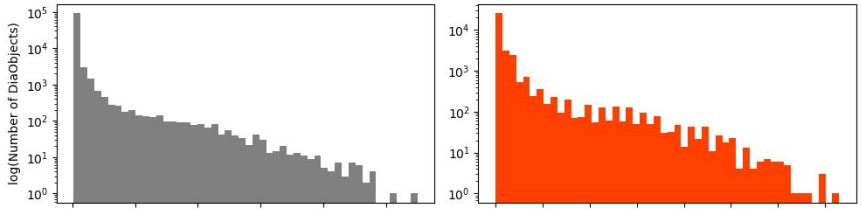
Notice how the distribution is peaked at small numbers of DiaSources (or r-band detections) -- these are time-variable sources that were only detected in a few difference images, and are most likely to have a faint time-variable flux (e.g., high-z supernovae that were only detectable for a short duration).

[6]:
```
fig, ax = plt.subplots(1, 2, figsize=(10, 3), sharey=False, sharex=False)

ax[0].hist(DiaObjs['nDiaSources'], bins=50, log=True, color='grey')
ax[0].set_xlabel('nDiaSources')
ax[0].set_ylabel('log(Number of DiaObjects)')

ax[1].hist(DiaObjs['rPSFluxNdata'], bins=50, log=True, color=plot_filter_colors['r'])
ax[1].set_xlabel('rPSFluxNdata')

plt.tight_layout()
plt.show()
```



1 7 LSST | Idle Recommended () (sciplat-lab:recommended) data-dev.lsst.eu Mode: Command Ln 1, Col 1 07a_DiaObject_Samples.ipynb

Funded by the European Union's Horizon 2020 - Grant N° 824064

Running CTA notebooks in the RSP

Work done by Sabine Elles with a CTA use case provided by Thomas Vuillaume

- Repackage the provided CTA container
  - Add libraries
  - Add some security (following rules provided by Rubin)
    - prevent notebook to run malicious code
  - Add JupyterLab version provided by Rubin
- Register the new docker container
- Declare it to the RSP (yaml file)

- Some difficulties with non-CentOS containers
  - Investigating the best way to overcome this

Funded by the European Union's Horizon 2020 - Grant N° 824064

Conclusions

- IMHO the Rubin Science Platform offers an interesting alternative approach to ESAP
- Open source
- Scalable
- Secure
- Will be maintained and developed through the next decade

- Containers provided by non Rubin projects can be relatively easily repackaged to run on the RSP

Funded by the European Union's
Horizon 2020 - Grant N° 824064