



# ESCAPE

European Science Cluster of Astronomy &  
Particle physics ESFRI research Infrastructures

## WP5 Status

John Swinbank — [swinbank@astron.nl](mailto:swinbank@astron.nl)



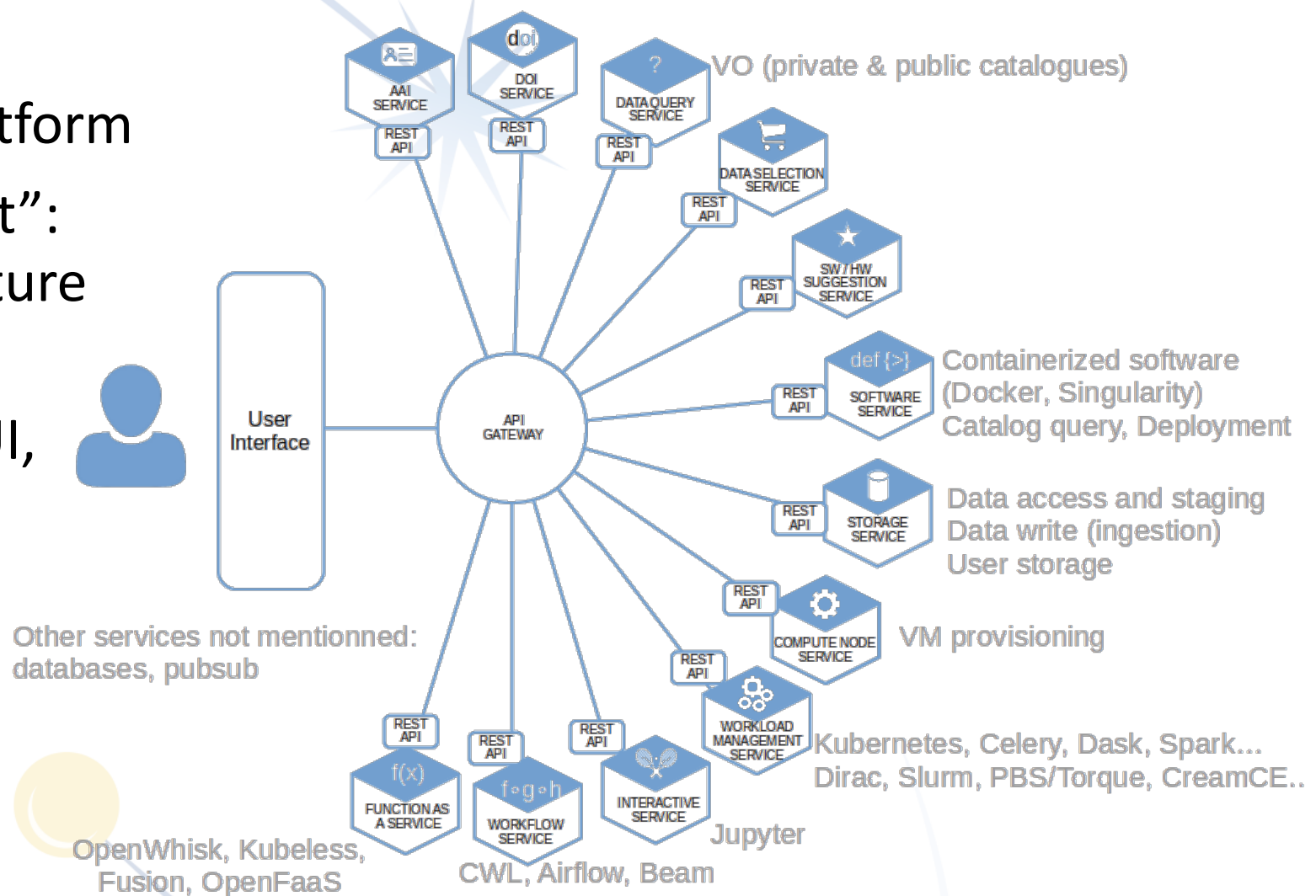
# Staffing & Admin

- As of late 2020:
- John Swinbank (ASTRON) took over from Michiel van Haarlem (also ASTRON) as WP5 coordinator.
- Zheng Meyer-Zhao (ASTRON), WP5 Tech Lead, is on indefinite leave of absence.
- Leadership transition provides a chance to evaluate current progress and assess goals for the year ahead...
- ...but also means we have lost some momentum and are gradually spooling back up.
- Opportunity for new technical leaders to step forward!



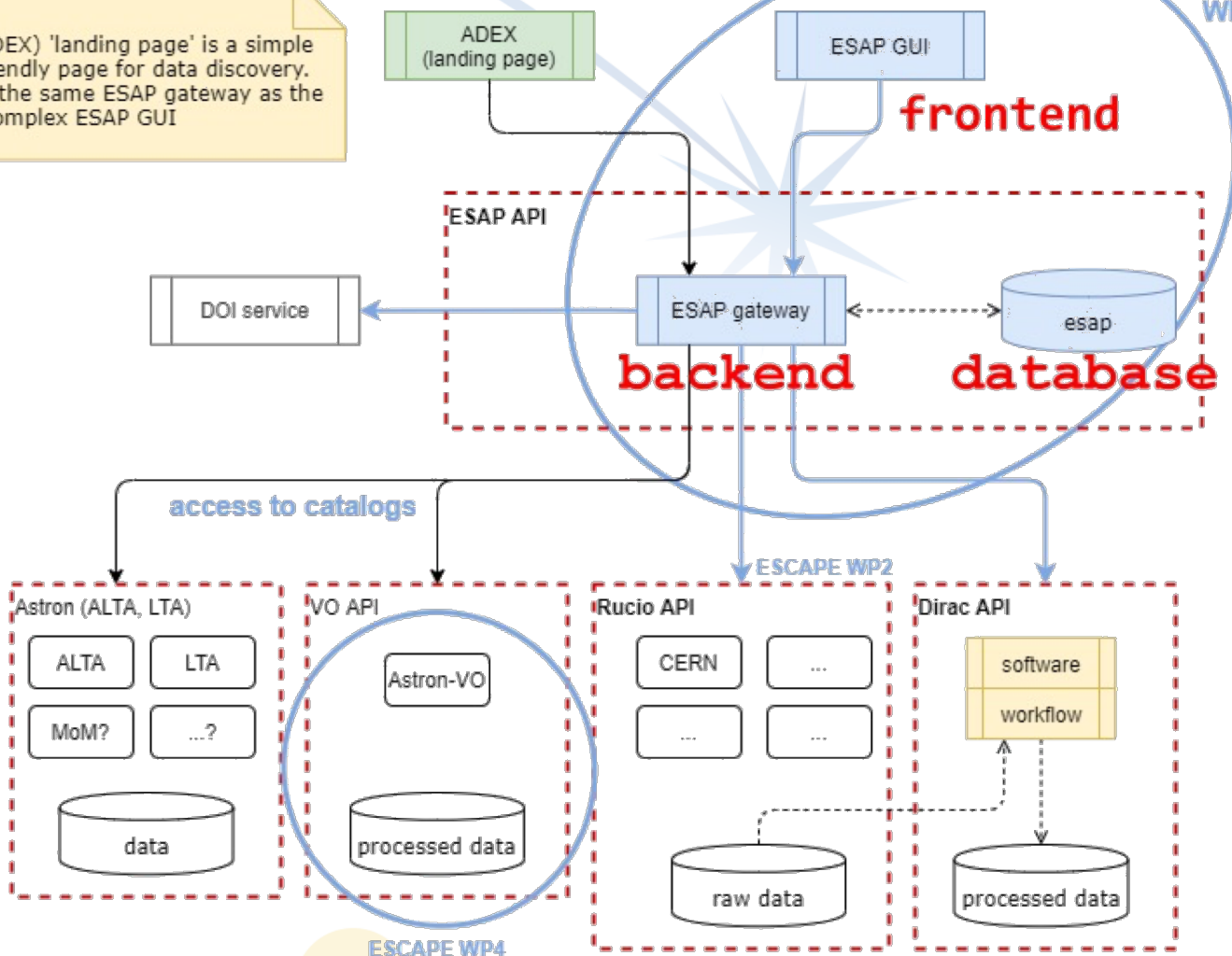
# ESAP: The hub in ESCAPE's wheel

- ESFRI Science Analysis Platform
- A “science platform toolkit”: bring your own infrastructure and service capabilities.
- Two part structure: web UI, and API gateway
- Focal point of a range of pluggable, independent microservices
- Designed to be robust & extensible



The (ADEX) 'landing page' is a simple user friendly page for data discovery. It uses the same ESAP gateway as the more complex ESAP GUI

ESCAPE WP5



All the information about archives, catalogs, datasets, available software and workflows, processing resources is described in the esap database.

**Rucio API** offers a standard way to access the catalogs of the data centers

**Dirac API** offers a standard way to access the processing resources of the data centers

Figure: Nico Vermaas (ASTRON)



# ESAP Queries in Action

The screenshot shows a web browser window with the URL `sdsc-dev.astron.nl/esap-gui/archives`. The page features a navigation bar with 'Archives', 'Query', 'Rucio', 'Interactive Analysis', and 'SAMP', along with a 'Login' button. The main content is organized into five columns, each representing a different data source:

- WSRT-Apertif:** Includes an image of a radio telescope and text describing Apertif Surveys. A button at the bottom reads 'Visit WSRT-Apertif Archives'.
- ASTRON VO:** Features the ASTRON Virtual Observatory logo and text explaining the standards for downloading astronomical data. A button at the bottom reads 'Visit ASTRON VO Archives'.
- LOFAR-LTA:** Shows a satellite view of the LOFAR radio telescope array and text about LOFAR LTA data. A button at the bottom reads 'Visit LOFAR-LTA Archives'.
- Zooniverse:** Displays the Zooniverse logo and text about the Classification Database. A button at the bottom reads 'Visit Zooniverse Archives'.
- Virtual Observatory (VO):** Shows the IVOA logo and text about the standards for downloading astronomical data. A button at the bottom reads 'Visit Virtual Observatory (VO) Archives'.



# ESAP Queries in Action

The screenshot displays the ESAP GUI interface for the WSRT-Apertif archive. The browser address bar shows the URL `sdc-dev.astron.nl/esap-gui/archives`. The page title is "Archive - WSRT-Apertif".

**Navigation:** Archives Query Rucio Interactive Analysis SAMP Login

**Instrument Details:**

Instrument	WSRT
Description	Apertif Surveys

**Apertif Surveys**

Data from the Apertif surveys include imaging and time-domain data. The time-domain products consist of high-time resolution filterbank data in the PSRFITS standard. The imaging data products include the raw observations in the measurement set (MS) standard format. In addition, processed data products are available, including calibration tables, calibrated visibilities, multi-frequency synthesis continuum images, polarization images and cubes, and uncleaned neutral hydrogen (HI) line and beam cubes. Full details of these data will be provided in upcoming papers (van Leeuwen et al. 2020, Adams et al. 2020).

[Visit WSRT-Apertif Archives](#)

**Data Retrieval**

Data can be retrieved via the Apertif Long Term Archive (ALTA). Public data can be downloaded from the ALTA web interface or retrieved via wget (necessary for MS files). Non-public data can be browsed with an ALTA account but cannot be retrieved via the ALTA web interface.

**Data Product Categories in WSRT-Apertif**

Data Product Category	Catalog name	Categories access
Imaging Survey raw data	WSRT-Apertif	<a href="#">Browse Catalog &amp; Run Queries</a>
Imaging Survey Processed Data	WSRT-Apertif	<a href="#">Browse Catalog &amp; Run Queries</a>

**Imaging Survey raw data**

In imaging mode, the Apertif correlator provides 24576 channels over a bandwidth of 300 MHz, which can be placed in the frequency range 1130-1750 MHz. The data is taken in full Stokes mode, and the combination of bandwidth and spectral resolution means that the Apertif imaging data are simultaneously sensitive continuum, polarization, and neutral hydrogen (H I) data.

The raw visibility data are automatically ingested into ALTA after completion of an observation. The visibilities are recorded in measurement-set (MS) format. With 24576 channels, 12 dishes, 40 beams, and 30 second integration, the size of a typical survey field (11.5 hour observation) is 4.6TB. Calibrator observations have 10 second integrations and vary from 4-5 minutes on source, resulting in a datasize of 83.2-104 GB per calibrator observation. Since a calibrator is placed in every beam, a total calibration scan (40 beams) results in 3.3-4.2 TB of data stored on ALTA. In order to minimize data volumes, the unnecessary data for each calibration observation (i.e., the 39 beams without a calibrator) are deleted from ALTA after the generation of inspection plots. This procedure is called "pruning" and results in a total size of 83.2-104 GB for each full calibrator scan. In addition to the raw data itself, inspection plots are generated for a quick look at the quality of the data.

**Imaging Survey Processed Data**

Imaging Survey data products are produced from running through the imaging pipeline, Apercal. This pipeline flagged the data, derived cross-calibration solutions, found self-calibration solutions from the target data, and produced multi-frequency synthesis continuum images and dirty non-continuum subtracted line cubes over the frequency range 1400-1400 MHz.



# ESAP Queries in Action

**ASTRON Data Explorer** Archives Query Rucio Interactive Analysis SAMP Login

**Apertif Data Collection Query**

Catalog: Apertif Target: [ ] RA (degrees): [ ] dec (degrees): [ ] search radius (degrees): [ ] Apertif Collections: Imaging

Processing Level: Raw DataProduct Type: All Data Release: APERTIF Imaging Data Release

**Submit**

**Query results for Apertif**

1 2 3 4 ... 319

Name	RA	Dec	fov	Dataset ID	Data Product Type	Data Product Subtype	Link to data
WSRTA190711129_B000.MS	202.8	30.5	0.2	190711129	visibility	uncalibratedVisibility	<a href="#">Download data</a> <a href="#">View Thumbnail</a>
WSRTA190711130_B001.MS	202.8	30.5	0.2	190711130	visibility	uncalibratedVisibility	<a href="#">Download data</a> <a href="#">View Thumbnail</a>
WSRTA190711131_B002.MS	202.8	30.5	0.2	190711131	visibility	uncalibratedVisibility	<a href="#">Download data</a> <a href="#">View Thumbnail</a>
WSRTA190711132_B003.MS	202.8	30.5	0.2	190711132	visibility	uncalibratedVisibility	<a href="#">Download data</a> <a href="#">View Thumbnail</a>
WSRTA190711133_B004.MS	202.8	30.5	0.2	190711133	visibility	uncalibratedVisibility	<a href="#">Download data</a> <a href="#">View Thumbnail</a>
WSRTA190711134_B005.MS	202.8	30.5	0.2	190711134	visibility	uncalibratedVisibility	<a href="#">Download data</a> <a href="#">View Thumbnail</a>
WSRTA190711135_B006.MS	202.8	30.5	0.2	190711135	visibility	uncalibratedVisibility	<a href="#">Download data</a> <a href="#">View Thumbnail</a>
WSRTA190711136_B007.MS	202.8	30.5	0.2	190711136	visibility	uncalibratedVisibility	<a href="#">Download data</a> <a href="#">View Thumbnail</a>



# ESAP: Simple Application Messaging Protocol (SAMP)

AstroView Observations Collections Parents AstroBase Admin Jobs: 0 SAMP Logout

### SAMP demo

register SAMP Ping unregister

target	ra	dec	beam	fov	central_frequency	n_channels	bandwidth	t_min	t_max	duration	task_id	obsid	dataprodut_type	dataprodut_subtype	product_key	equinox
S2149+3848	-34.1918	39.8101	33	0.5	1361.25	12288	150	2020-02-29T04:51:37	2020-02-29T16:21:37	11.5	200229034	200229034	visibility	calibratedVisibility	S2149+3848.UVFITS	J2000
S2149+3848	-33.5829	39.8173	34	0.5	1361.25	12288	150	2020-02-29T04:51:37	2020-02-29T16:21:37	11.5	200229034	200229034	visibility	calibratedVisibility	S2149+3848.UVFITS	J2000
S2207+4114	-28.7058	41.4426	23	0.5	1361.25	12288	150	2019-10-	2019-10-	11.5003	191005041	191005041	visibility	calibratedVisibility	S2207+4114.UVFITS	J2000
S2210+4340	-27.9715	42.6412	1	0.5	1361.25	12288	150	2020-02-29T04:51:37	2020-02-29T16:21:37	11.5	200229034	200229034	visibility	calibratedVisibility	S2210+4340.UVFITS	J2000
S2201+4340	-30.4913	42.657	3	0.5	1361.25	12288	150	2020-02-29T04:51:37	2020-02-29T16:21:37	11.5	200229034	200229034	visibility	calibratedVisibility	S2201+4340.UVFITS	J2000
S2201+4340	-30.1763	43.0641	10	0.5	1361.25	12288	150	2020-02-29T04:51:37	2020-02-29T16:21:37	11.5	200229034	200229034	visibility	calibratedVisibility	S2201+4340.UVFITS	J2000

Table Access Protocol (TAP) Query

Table List:

- TAP\_1\_TAP\_SCHEMA.tables
- TAP\_2\_apertif\_dr1.continuum
- TAP\_3\_apertif\_dr1.calibrated
- TAP\_4\_apertif\_dr1.calibrated

Current Table Properties:

Label: TAP\_4\_apertif\_dr1.calibrated\_visibilities  
 Location: TAP\_4\_apertif\_dr1.calibrated\_visibilities  
 Name: calibrated\_visibilities  
 Rows: 1,000  
 Columns: 20  
 Sort Order: [arrow]  
 Row Subset: All  
 Activation Actions: 1 / 8

Service Capabilities:

Query Language: ADQL-2.0 Max Rows: 20000 (default) Uploads: 20Mb

ADQL Text:

```
Mode: Synchronous
1
SELECT TOP 1000 * FROM aperitif_dr1.calibrated_visibilities
```

Table Browser for 4: TAP\_4\_apertif\_dr1.calibrated\_visibilities

	target	ra	dec	beam	fov	central_frequency	n_chann...	bandwidth
1	S2149+3848	-34.1918	39.8101	33	0.5	1361.25	12288	150.
2	S2149+3848	-33.5829	39.8173	34	0.5	1361.25	12288	150.
3	S2207+4114	-28.7058	41.4426	23	0.5	1361.25	12288	150.
4	S2207+4114	-28.0814	41.4434	24	0.5	1361.25	12288	150.
5	S2207+4114	-27.7663	41.8479	30	0.5	1361.25	12288	150.
6	S2207+4114	-28.3946	41.8487	29	0.5	1361.25	12288	150.
7	S2207+4114	-28.0794	42.254	36	0.5	1361.25	12288	150.
8	S2216+4340	-27.9715	42.6412	1	0.5	1361.25	12288	150.
9	S2201+4340	-30.4913	42.657	3	0.5	1361.25	12288	150.
10	S2201+4340	-30.1763	43.0641	10	0.5	1361.25	12288	150.
11	S2201+4340	-28.8842	43.8729	25	0.5	1361.25	12288	150.
12	S2201+4340	-29.2055	44.28	30	0.5	1361.25	12288	150.
13	S2201+4340	-31.1664	44.2716	27	0.5	1361.25	12288	150.
14	S2201+4340	-28.8593	44.2809	29	0.5	1361.25	12288	150.

Total: 1,000 Visible: 1,000 Selected: 0





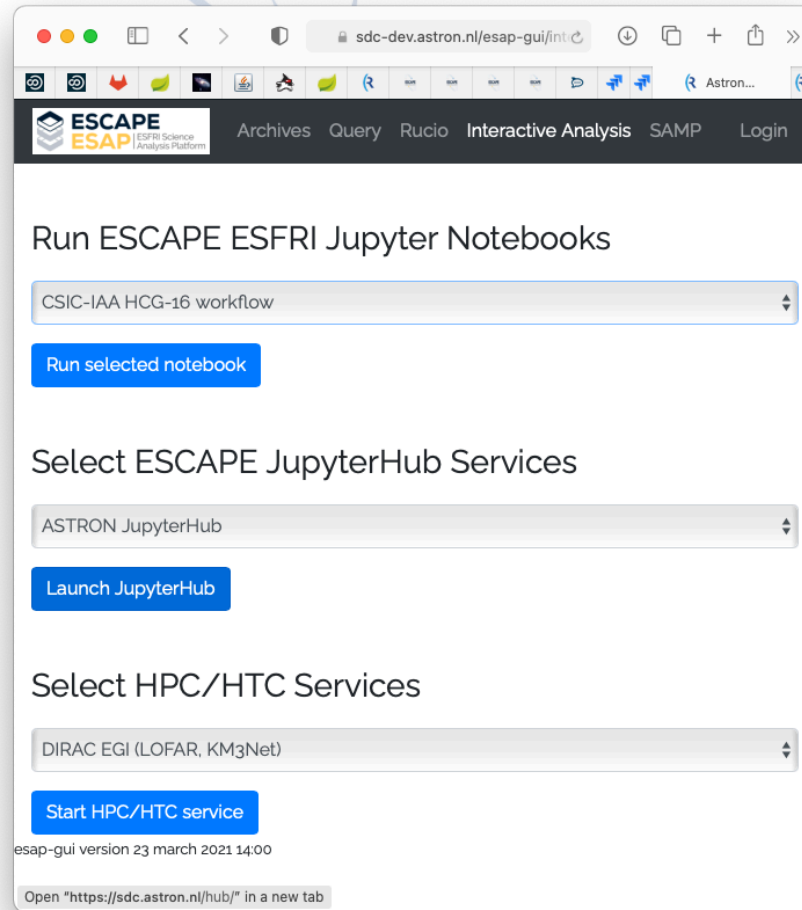
# ESAP: Rucio

- ESAP — Rucio is currently a demo implementation by Hugh Dickinson
- which is hard to demo because AAI depends on X509 which is used to get a token. OIDC would make things easier
- Also Rucio REST interaction runs under non-standard port, which is blocked by default at ASTRON where ESAP development version is running
- Implementation is fairly simple: list files in specific scope
- Searching for data (other than scope) is not really implemented



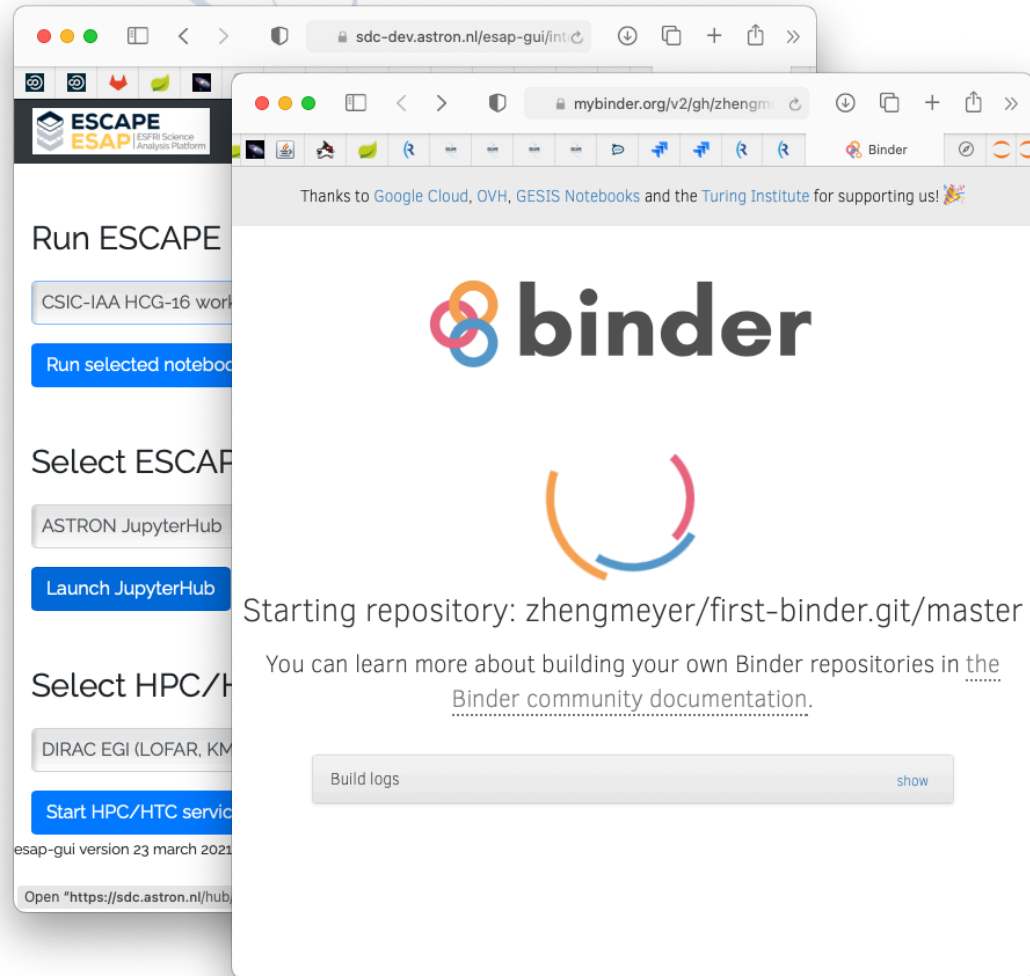
# ESAP: Interactive Data Analysis & HPC/HTC Services

- Currently, ESAP acts as a “hub”, which effectively links out to JupyterHub & H[TP]C facilities.
- Notebooks running on e.g. Binder can access data made available through the VO and other standard interfaces.



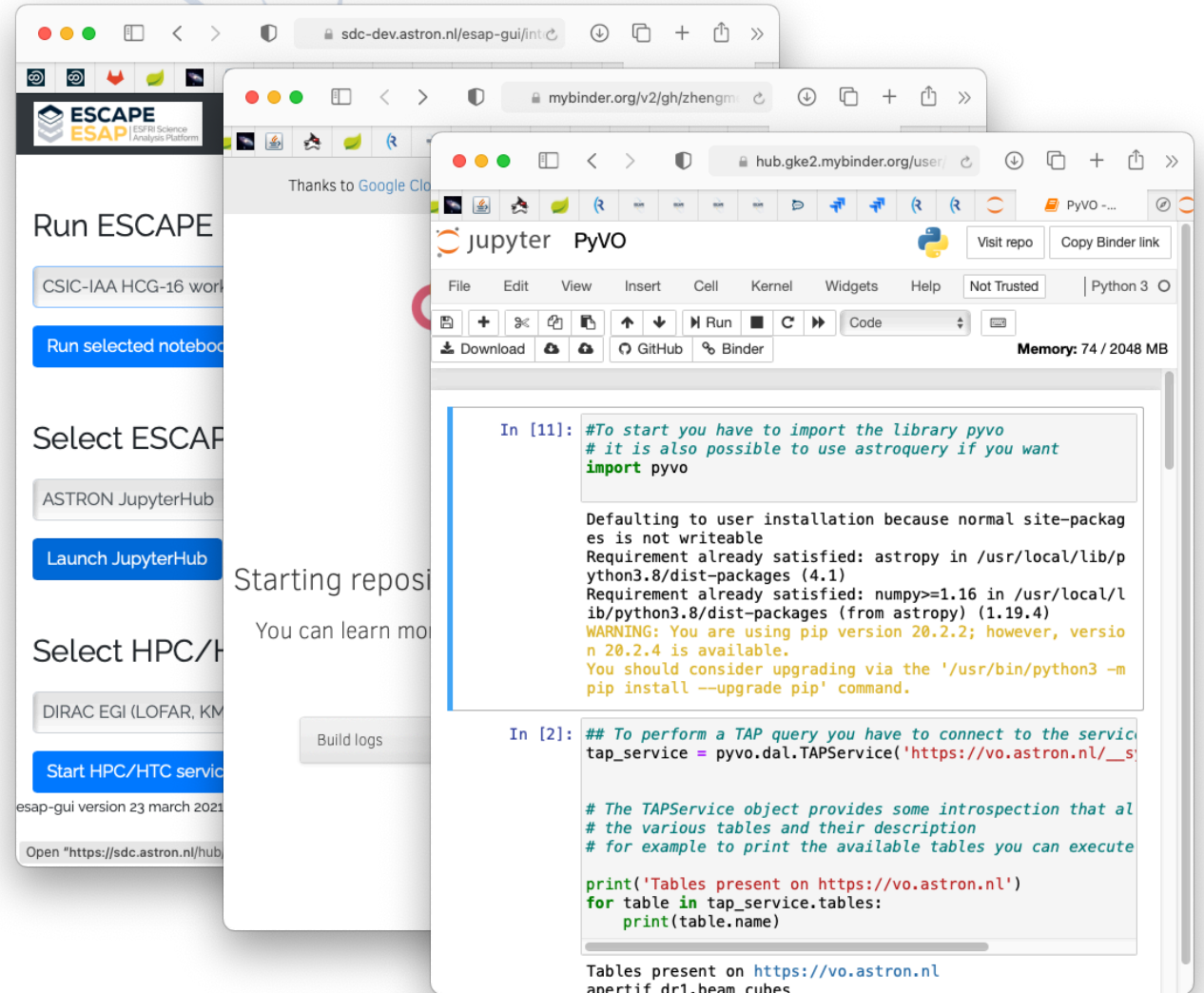
# ESAP: Interactive Data Analysis & HPC/HTC Services

- Currently, ESAP acts as a “hub”, which effectively links out to JupyterHub & H[TP]C facilities.
- Notebooks running on e.g. Binder can access data made available through the VO and other standard interfaces.



# ESAP: Interactive Data Analysis & HPC/HTC Services

- Currently, ESAP acts as a “hub”, which effectively links out to JupyterHub & H[TP]C facilities.
- Notebooks running on e.g. Binder can access data made available through the VO and other standard interfaces.



The image shows a composite of three browser windows. The leftmost window is the ESAP GUI, displaying options to 'Run ESCAPE' (with a 'Run selected notebook' button), 'Select ESCAPE' (with 'ASTRON JupyterHub' and a 'Launch JupyterHub' button), and 'Select HPC/HTC' (with 'DIRAC EGI (LOFAR, KM)' and a 'Start HPC/HTC service' button). The middle window shows a 'Starting repository' message. The rightmost window is a Jupyter Notebook running on PyVO, showing the execution of two code cells. The first cell imports the 'pyvo' library, and the second cell uses 'pyvo.dal.TAPService' to connect to a service and print available tables.

```

In [11]: #To start you have to import the library pyvo
# it is also possible to use astroquery if you want
import pyvo

Defaulting to user installation because normal site-packages
is not writeable
Requirement already satisfied: astropy in /usr/local/lib/p
ython3.8/dist-packages (4.1)
Requirement already satisfied: numpy>=1.16 in /usr/local/l
ib/python3.8/dist-packages (from astropy) (1.19.4)
WARNING: You are using pip version 20.2.2; however, versio
n 20.2.4 is available.
You should consider upgrading via the '/usr/bin/python3 -m
pip install --upgrade pip' command.

In [2]: ## To perform a TAP query you have to connect to the service
tap_service = pyvo.dal.TAPService('https://vo.astron.nl/_s

# The TAPService object provides some introspection that al
# the various tables and their description
# for example to print the available tables you can execute

print('Tables present on https://vo.astron.nl')
for table in tap_service.tables:
    print(table.name)

Tables present on https://vo.astron.nl
anertif dr1.beam.cubes
    
```



# ESAP: Future Interactive Data Analysis

- Lift that linking to JupyterHub facilities to the next level.
- That is, instead of simply linking to an existing notebook, the user should be able to:
  - Select data in the ESAP interface
  - Select notebook contents from WP3/OSSR
  - Search for an appropriate notebook service
  - Stage data & software to the notebook service
  - Launch an interactive analysis session
- Jupyter seems like the 800 lb gorilla in this conversation, but do we need other services too?



# ESAP: Future H[TP]C

- Effectively the same goals as for interactive analysis/Jupyter.
- But there's much more diversity of target systems; unclear if there's a Jupyter-like gorilla that we can target in this space.
- DIRAC seems to be the name that's frequently mentioned, but I've not seen a real analysis of whether that's sufficient.
- Seems like there's a lot of work to do here to really understand our scope, use cases, etc.



# Rubin Observatory Science Platform

- Work carried out at CNRS-LAPP to investigate the Vera C. Rubin Observatory Science Platform.
- RSP provides tight coupling of a search/discovery portal, Jupyter notebook service, and VO APIs.
- A different regime from ESAP, but an opportunity to learn.

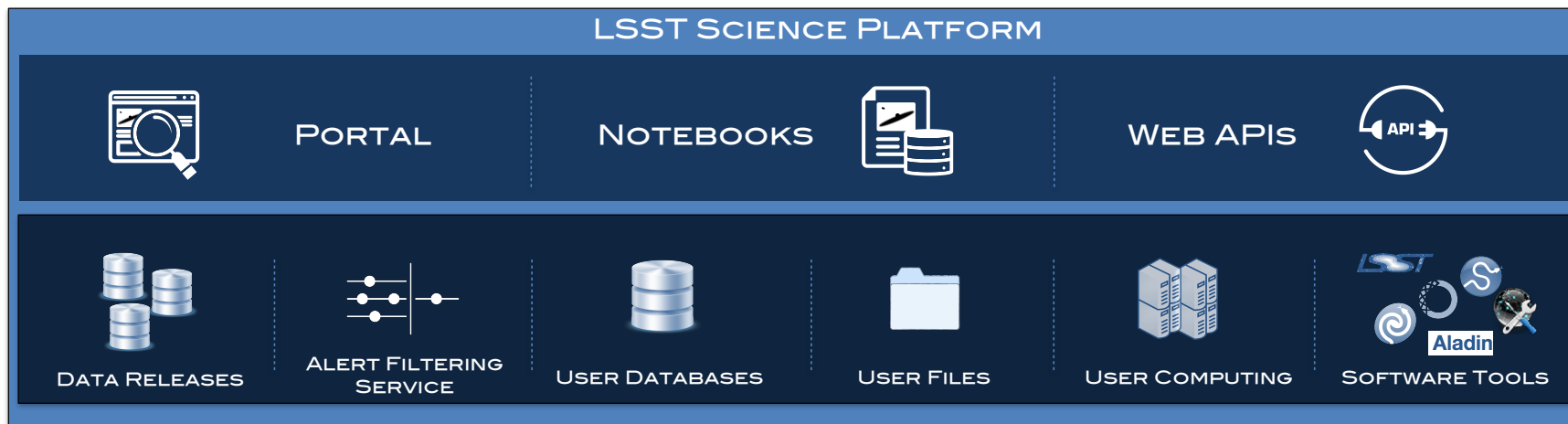


Figure: Vera C. Rubin Observatory



# Other WGs

- At this stage in ESCAPE development, it's clear that we need to think of a unified project: ESAP is only useful if it integrates with the products of other work packages.
- Hi, WP2! 🙌
- Also regular meetings with WP3 (Software & Service Repository).
- Metadata “task force” focusing on Jupyter notebooks just started.
- Starting with a use case based on existing notebook analysis
  - <https://github.com/AMIGA-IAA/hcg-16>
  - <https://indico.in2p3.fr/event/23933/>





# Milestones

- MS31: Second WP5 workshop to analyse prototype performance
  - February 2021 (delayed due to leadership change)
- MS32: Integration of Science Platform with OSSR repository
  - August 2021
- MS33: Integration of Science Platform with Data Lake
  - October 2021



# WP5 Next Steps

- Focus on concrete & credible development plan driven by use cases.
- Develop core technical team to clarify scope.
- Publish ESAP development roadmap.

