

# Astronomy in Galaxy

Working with Astro Data, Big Archives, Tool catalogs

Volodymyr Savchenko (EPFL)  
Francois Morier-Genoud (EPFL) Denys Savchenko (UPC)  
Andrii Neronov (EPFL, UPC)

European Galaxy Days  
Freiburg, 5 Oct 2023

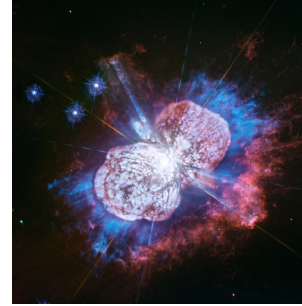
# Astronomy goes Multi-Messenger

Last decade key **new kinds of emission** were discovered, while conventional telescopes dramatically upgraded.

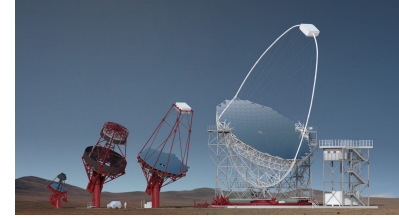
**Number of alerts** and **volume of data** we deal with **increased by couple orders of magnitude in the last years**, and several nearly-ready telescopes promise another comparable increase

Combining these data quickly is difficult.

“Just” a star



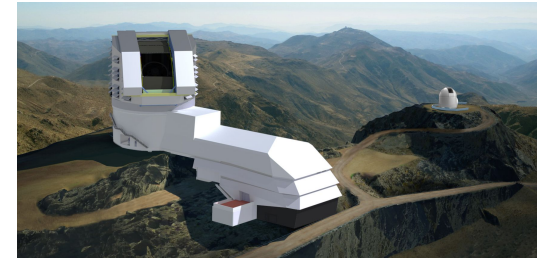
Gamma-ray



Radio

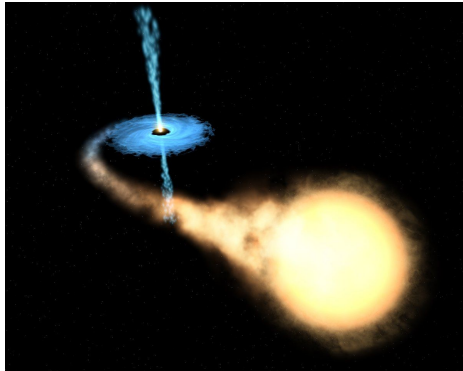


Visible

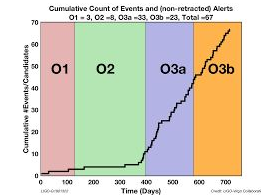


Gravity

Star and black hole

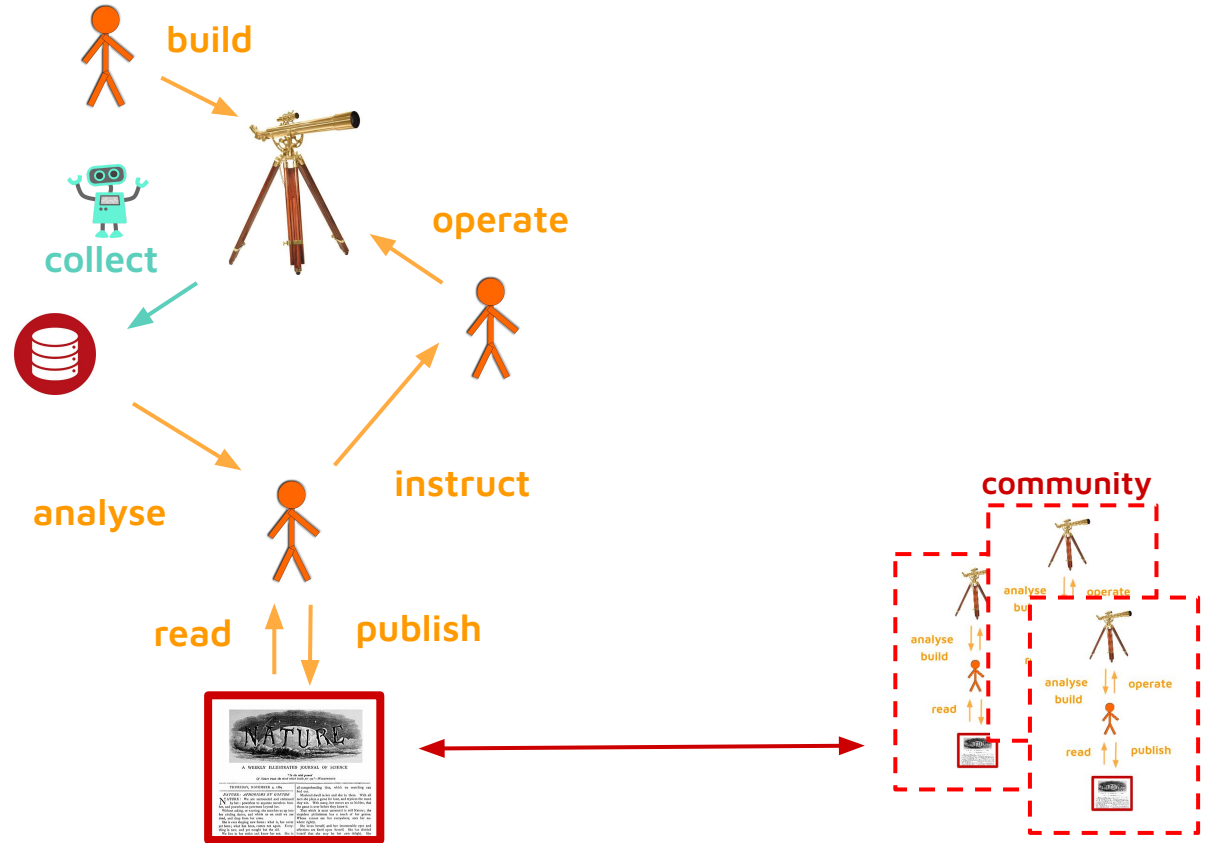


Two neutron stars



# Mostly-human Astronomy

- Reaction to sky: **slow**
- Reaction to papers: **slow**
- Trials (p-hacking): **uncontrolled**
- Publishing: **slow**
- Scalability: **bad**
- Creativity: **high**
- Communication: **nuanced but imprecise and slow**

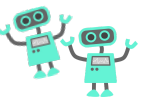
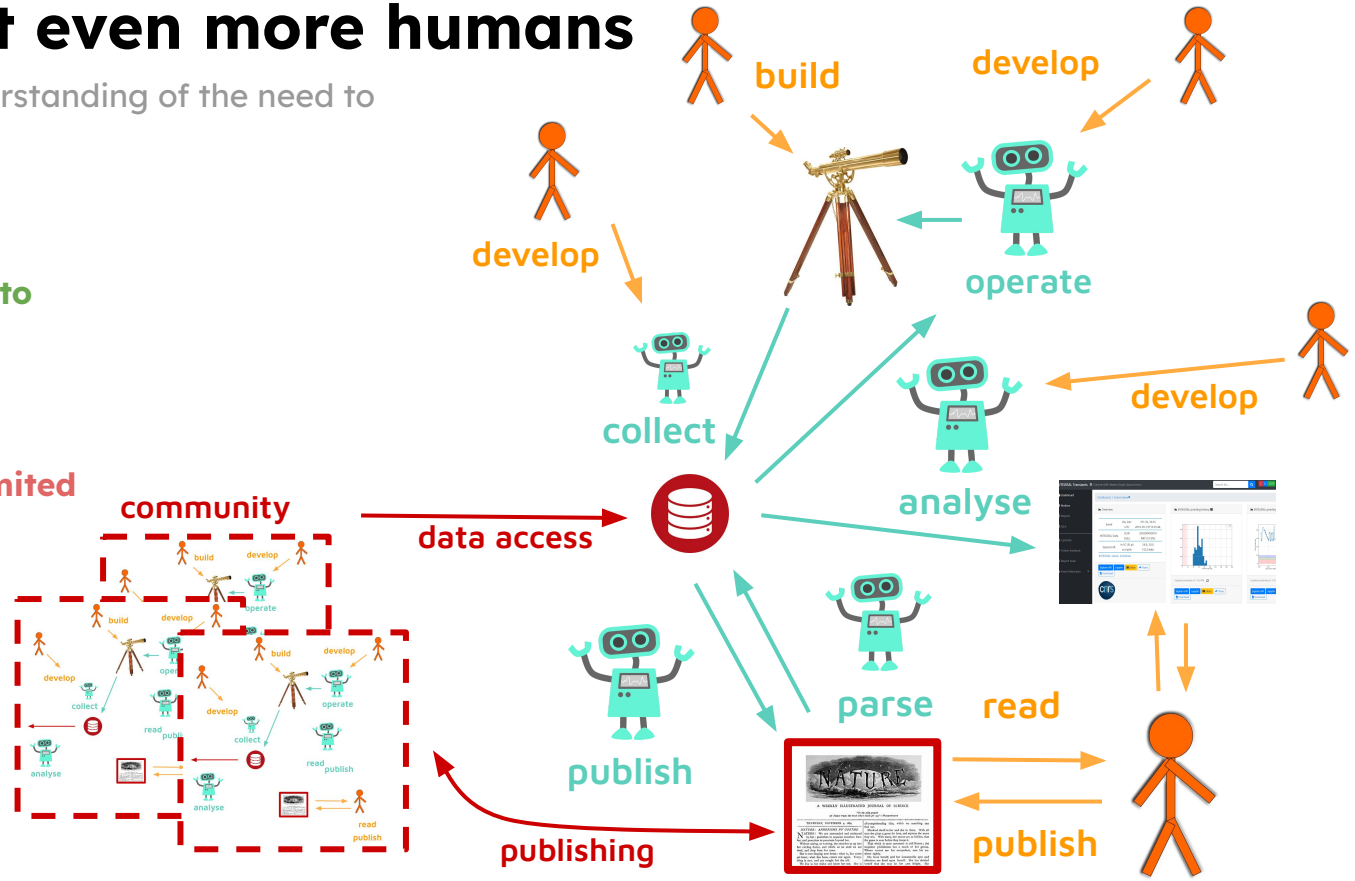


Human reaction and processing **is slow**, even if it's within even one person. But people are **adaptable**

# More robots, but even more humans

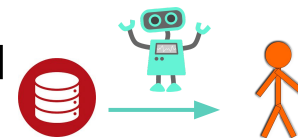
Community developed an understanding of the need to robotize

- Reaction to sky: **fast**
- Reaction to literature: **fast**
- Trials (p-hacking): **possible to control**
- Publishing: **fast**
- Scaling: **good**
- Creativity: **low**
- Communication: **precise, limited**

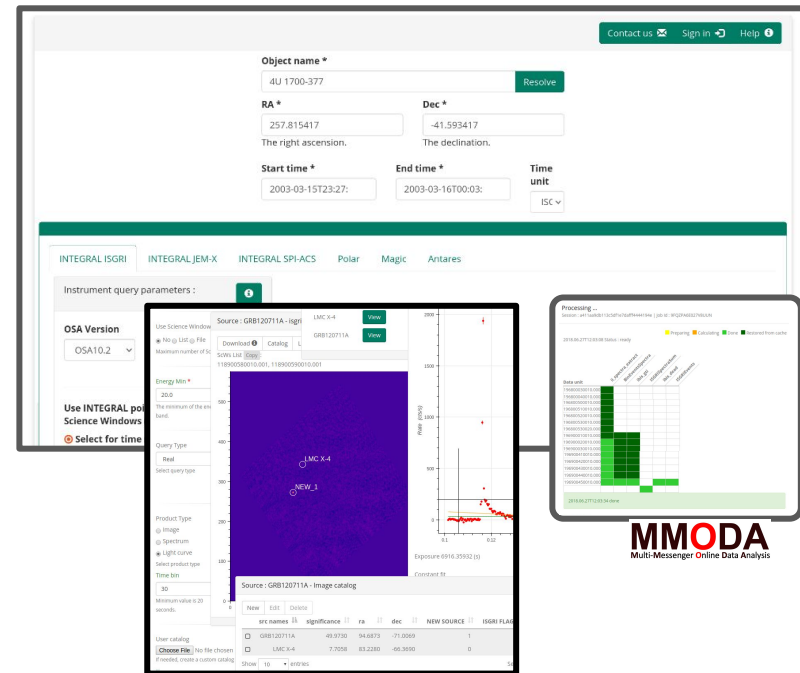
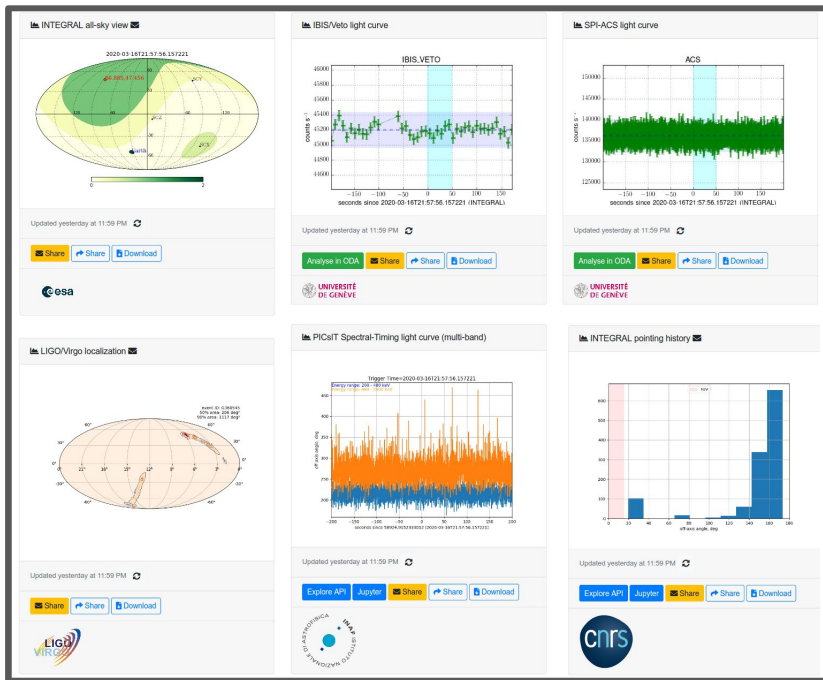


- **Making smart robots is hard:** always lacking developers who are also research scientists.
- If all is automated, **scientists have hard time seeing what's going on**, since they do not speak robot
- Robots are **fast**, but **lack creative reaction** in new situations.

# Tools for exploring, transforming research data



Various tools were developed in the Astro community. Tools are developed by large telescope collaborations or agencies.



**Hard to build these tools**, need expert astronomers with state-of-the-art tool-building skills self.

<https://marketplace.eosc-portal.eu/services/astronomical-online-data-analysis-astroda>

<https://github.com/oda-hub>

# Galaxy Astronomy: **FITS** file support, preview with **AladinLite**

We took example of HDF5 to extend Galaxy with **FITS** format: **identification**, **parsing**, conversion to **tabular** to use **existing Galaxy tools**.

**Visualization** is based on **AladinLite JavaScript application**. Sky images have stable reference with microarcsecond accuracy (equivalent to 0.03mm on Earth surface)

Allows to navigate the image, explore **external data sources**, source **catalogs**, adjust presentation.

Disadvantage of **JS UI** is that what is done in it does not make **a reproducible workflow**. So we also made **some tools** to replicate of the functionalities. These and other tools are meant to be used on usegalaxy.eu

The screenshot displays the Galaxy Astronomy web interface. The main window shows a FITS image viewer titled "FITS aladin viewer: legacysurvey\_image.fits". The image is a color-coded astronomical image of a galaxy, with a central bright region and a surrounding diffuse structure. The image is overlaid with a green grid. The viewer includes a toolbar with various icons for zooming, panning, and other navigation functions. The top navigation bar shows "Galaxy Europe" and "Workflow". The left sidebar contains a "Tools" section with a search bar and a list of tool categories. The right sidebar shows a "History" section with a search bar and a list of uploaded files, including "1. legacysurvey\_image.fits" (1015.3 KB) and "0 PrimaryHDU IMAGE (360, 360)". Below the history section, there are several tool cards, including "astropy fits2bitmap", "astropy fits2csv", and "astropy fitsinfo". The bottom right corner of the interface shows a "WORKFLOWS" section with a list of all workflows.

Most of the work by *Francois Morier-Genoud* <https://galaxyproject.org/news/2023-06-20-esg-wp5-astronomy-fits/> [Video](#)

# Adding Astro Data to Galaxy: **I**VOA archives

Archives datasets are measured in Pb, and current telescopes reach “**exoscale**” - “**BigData**”, motivating **bringing code to the data** in **web-based** platforms linked to **HPC**. Particle physics developed **Grid** and some **key technologies derive from it** (e.g. OakRidge team support of **Rucio**, see PASC [talk](#); also **ARC**)

Astronomy has interoperability standards for describing archives, developed International Virtual Observatory Alliance (**I**VOA) but also **ESA, NASA, ESO**.

We added first interface to query **I**VOA **TAP (Table Access Protocol)** archives from Galaxy ([demo video](#)).

Now working on more interactive UI, facing questions about how to operate Galaxy API from the UI

**Astronomical Archives (IWOA)** queries astronomical archives through Virtual Observatory protocols (Galaxy Version 0.9.0)

**Tool Parameters**

**Archive Selection**

Query specific IWOA TAP archive

**Archive \***

The VO @ ASTRON TAP service

**Query selection**

**ADQL Query Selection**

IWOA obscure table query builder

**Observation target name** - optional

**Galaxy** | Workflow | Visualize | Shared Data | Admin | Help | User | Using 4.2 GB

**RESOURCES PREVIEW ARCHIVE:** <https://vo.astron.nl/tap>

ADQL query: SELECT TOP 100 \* FROM iwoa.obscure

dataproduct_type	dataproduct_subtype	calib_level	obs_collection	obs_id	obs_title
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B001_0
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B001_1
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B001_2
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B001_3
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B008_0
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B008_1
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B008_2
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B008_3
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B015_0
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B015_1

History

Unnamed history

Most of the work by *Francois Morier-Genoud*

# Astrophysical BigData FAIR Workflows Reproducibility

We are especially concerned with publishing of FAIR BigData HPC workflows since they depend on sometimes **unique** and potentially **volatile external storage and compute resources**.

Reproducing the workflow like that is only possible on **large specialized HPC**, and is associated with non-negligible **energy costs**

We developed **runtime introspection** intercepting **external resource queries** to create annotations.

Another challenge is tracking **provenance of very large workflows**, often with many similar fragments.

```
Appendix-\ref{sec:association}).}

\begin{figure}
\centering
\includegraphics[width=1.\linewidth]{acs_lc}
\caption{SPI-ACS light curve of GRB-170817A (100-ms time resolution), detected 2 seconds after GW170817. The red line highlights the 100-ms pulse, which has an S/N of 4.6. The blue shaded region corresponds to a range of one standard deviation of the background.}
\label{fig:acs_lc}
\end{figure}
```

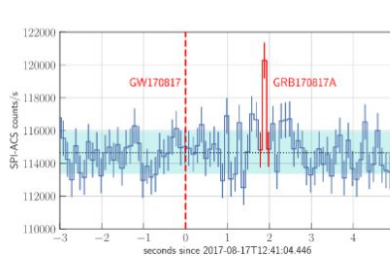


FIG. 2.— SPI-ACS light curve of GRB 170817A (100 ms time resolution), detected 2 seconds after GW170817. The red line highlights the 100 ms pulse, which has an S/N of 4.6. The blue shaded region corresponds to a range of one standard deviation of the background.

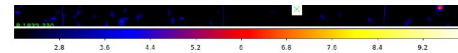


Fig. 12. Zoom of the significance map of the sky region around GX 5-1 in the 20–30 keV energy band obtained from a selection of 44 science windows belonging to the GPS and GCDE programs carried out in 2013 with a pointing offset of less than three degrees from GX 5-1. This image can be obtained from the following URL (ROCrated, validated at 2023-01-12 14:28 ).

spectra found in examples makes it possible to generate the results described below online.

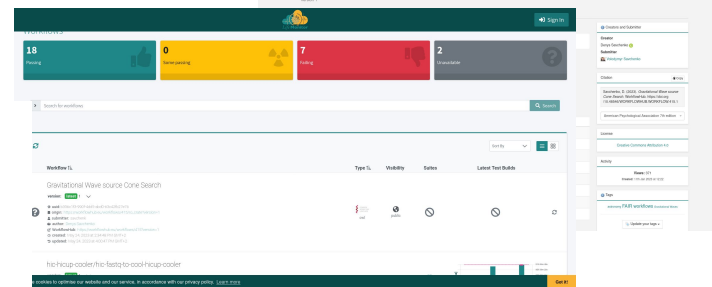
At the first stage of analysis, we determine the bright sources in the source field. We generate a mosaic image of the field and

$T_{in}$
norm <sub>disk</sub>
$kT$
$\tau$
norm
factor
$\chi^2/d.o.f.$
Flux (5–20)
Flux (20–100)

Notes. <sup>(a)</sup> Fluxes in the source field. We generate a mosaic image of the field and we report

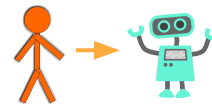
LifeMonitor

WorkflowHub





# Development space: help scientists make tools

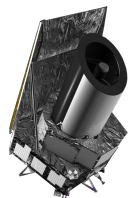
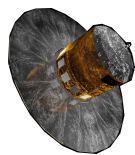
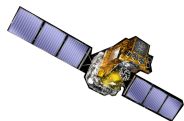


There are much more scientists who can make a **jupyter notebook** than write organized code.

Notebooks were already developed with numerous **JupyterHubs, Google-collab, ESA DataLabs, RenkuLab** etc

We add parameterized (papermill-style) notebooks in **semantic annotation ontology** terms for scientists and robots reuse

This process creates a collection of tools for publication-ready analysis of Gaia, HESS, Euclid, INTEGRAL, LIGO ... (to be published soon)



The screenshot shows a Jupyter notebook with the following code:

```
[2]: # name_input = "PKS 1156+295"
# name_input = "Mrk 421"
# name_input = "Crab"
# name_input = "PKS 2155-304"
# name_input = "OSO B0229+200"
# name_input = "NGC 1068"
name_input = ""
ra_s = 166.113808 # ra in degree
dec_s = 38.208833 # dec in degree
image_size = 3. # in arcmin (integrate flux of all sources within this radius)
radius_photometry = 1. # in arcsec
dr=9 # data release
image_bands='g' #
pixsize=1. # arcsec per pixel

[4]: from astropy.io import ascii
import numpy as np
from numpy import pi, cos, sin
import os

# astroquery API that will be traced by Renku plugin
from astroquery.ned import Ned
from astroquery.simbad import Simbad
```

The plot shows Flux vs Time, MJD, with a legend indicating parameters: 0.1 s, 1 s, SN = 8.1, 10 s, SN = 10.6. A red arrow points to the plot with the label "Lightcurve PNG".

Another red arrow points to the code cell with the label "Actual Lightcurve data from oda-api (no visualizer in galaxy yet)".

Most of the work for ESG by *Denys Savchenko*

<https://odahub.io/docs/guide-development/>

# Galaxy in Astronomy: Future, Impact, Synergies

We focus on integrating interoperable **components, integrations, standards**, to bring Galaxy closer to frameworks of Astro community and **reduce cost of adoption**, including:

- Visualization plugins for data types
- Astro tool catalogs
- Data catalogs (IVOA, etc)
- Compute (e.g. ARC) and storage (Rucio) resources

We are working with Astro research infrastructures, telescopes, and the community to **explore possibilities for adopting Galaxy as part of the major astrophysical** research environments.

It seems quite attractive to provide a single entrypoint like [usegalaxy.eu](http://usegalaxy.eu) at least for some common astronomical tools.

Integration with **large HPC and BigData** is crucial for Astro adoption, and **EuroScienceGateway** is making progress in this direction.

**EU/EOSC ESCAPE** project developed [ESAP](#) platform joining particle physics and astronomy, **ESA** developed [DataLabs](#), UNIGE Developed [DACE](#) and [MMODA](#), all provide similar many features in different contexts. These platforms are powerful, diverse, hard to discover.

The image displays a collage of screenshots from various astronomical data platforms. At the top left is the ESCAPE ESAP interface, showing a search for 'h.e.s.s.' in the 'ESAP IVOA Query' section. Below it is a 'List of VO Resources' table. To the right is a terminal window showing command-line output. In the center and bottom are screenshots of the 'DataLabs' and 'Datalab Launch' interfaces. A circular diagram with four green boxes labeled 'Find', 'Launch', 'Switch', and 'Use' is overlaid on the screenshots, indicating a workflow. The 'DataLabs' interface shows various tool categories like 'ESDC Analysis Lims', 'CRDS Desktop Analysis', and 'CRDS Command-line'. The 'Datalab Launch' interface shows a 'Name Options' field and a 'Launch' button. The bottom right screenshot shows a dark-themed interface with a grid of data points. At the bottom of the collage is a row of European Union member state flags and the text 'THE EUROPEAN SPACE AGENCY'.



The End

