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

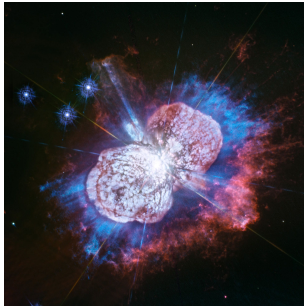
Astronomy goes Multi-Messenger

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

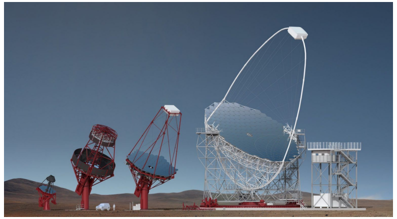
Number of alerts and **volume of data** we deal with **increased by couple orders of magnitude in the last years**, and several nearly-ready telescopes promise another comparable increase

Combining these data quickly is difficult.

"Just" a star

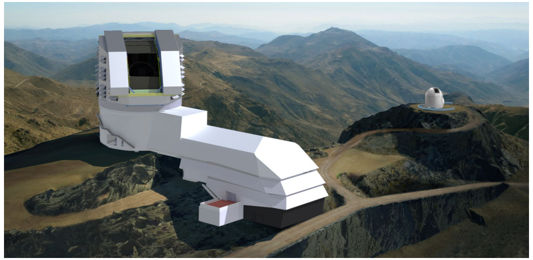


Gamma-ray

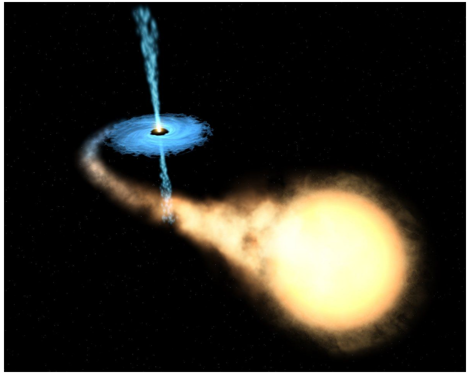


Radio

Visible

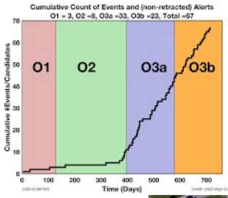


Star and black hole



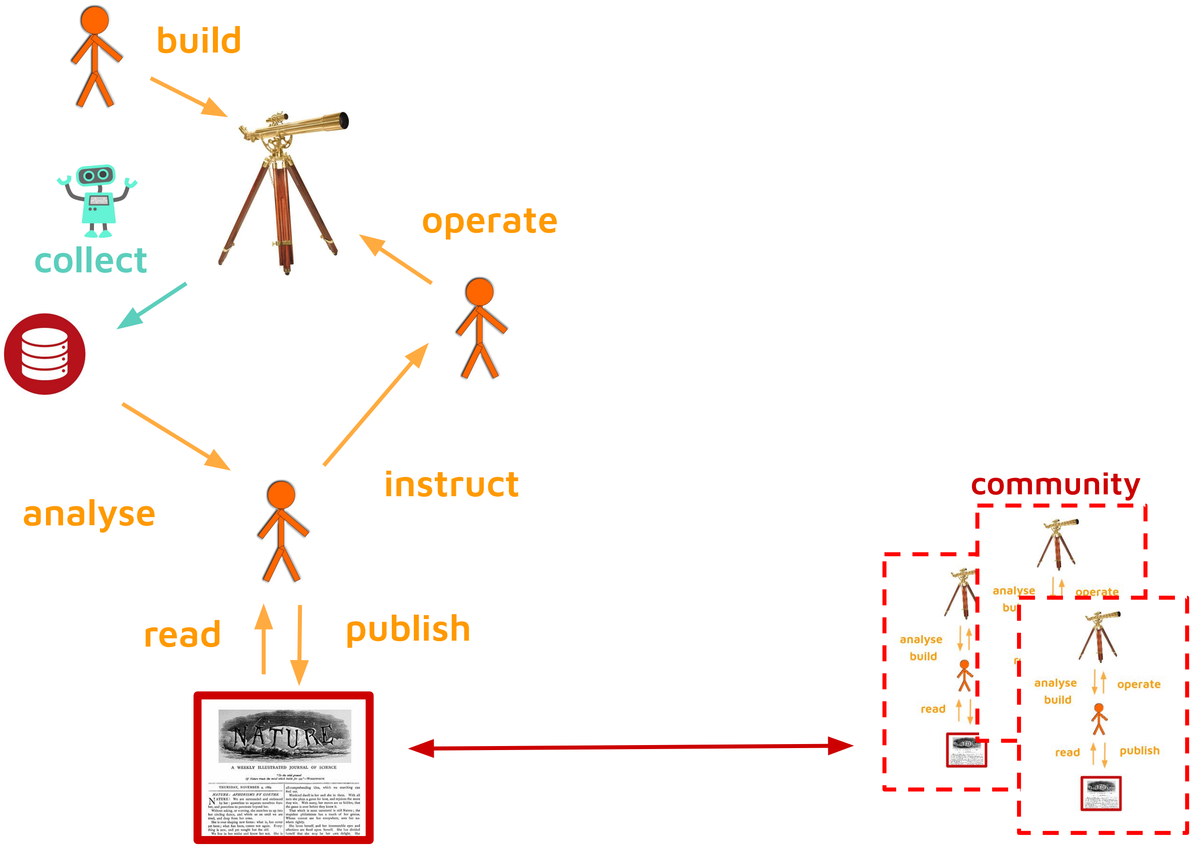
Two neutron stars

Gravity



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**

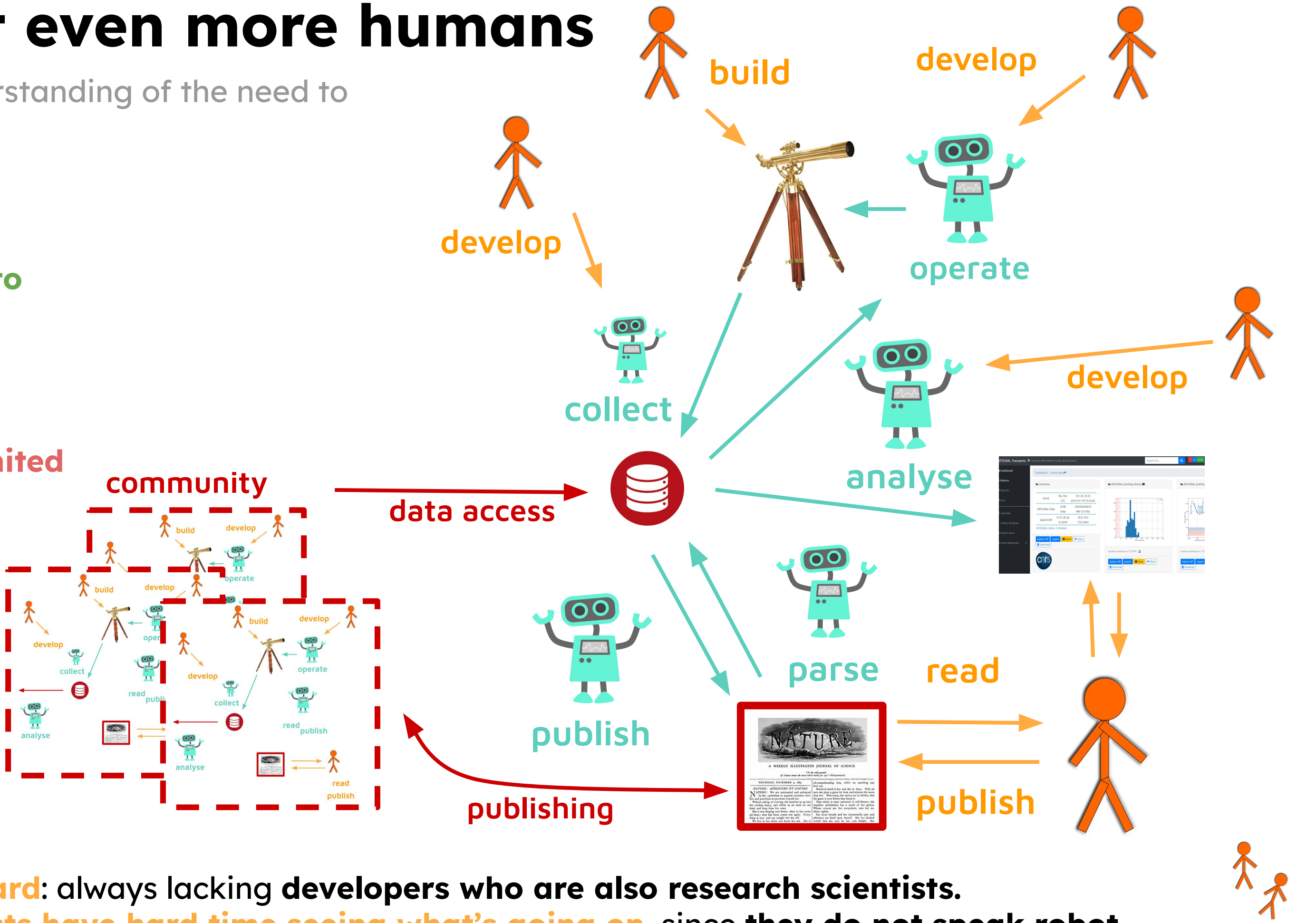


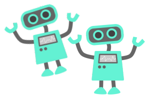
Human reaction and processing **is slow**, even if it's within even one person. But people are **adaptible**

More robots, but even more humans

Community developed an understanding of the need to robotize

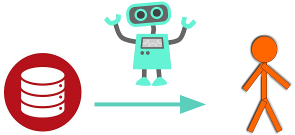
- Reaction to sky: **fast**
- Reaction to literature: **fast**
- Trials (p-hacking): **possible to control**
- Publishing: **fast**
- Scaling: **good**
- Creativity: **low**
- Communication: **precise, limited**



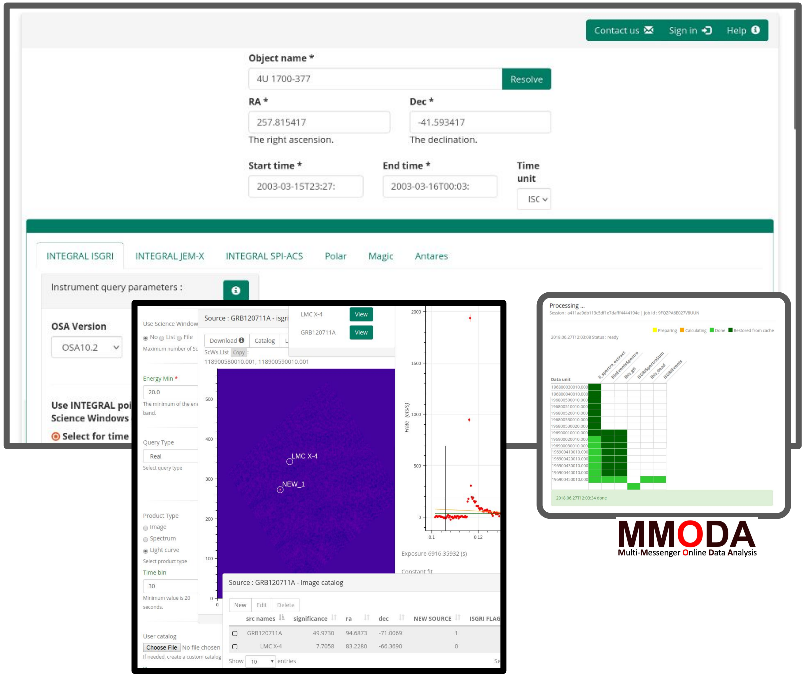


- **Making smart robots is hard**: always lacking **developers who are also research scientists.**
- If all is automated, **scientists have hard time seeing what's going on**, since **they do not speak robot**
- Robots are **fast**, but **lack creative reaction** in **new situations.**

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-astrooda
https://github.com/oda-hub

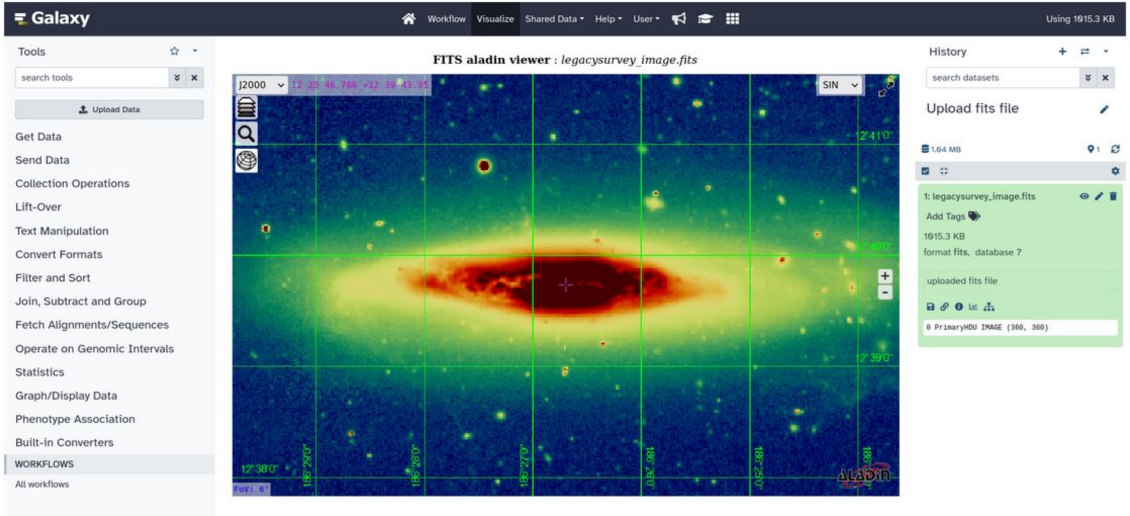
Galaxy Astronomy: FITS file support, preview with AladinLite

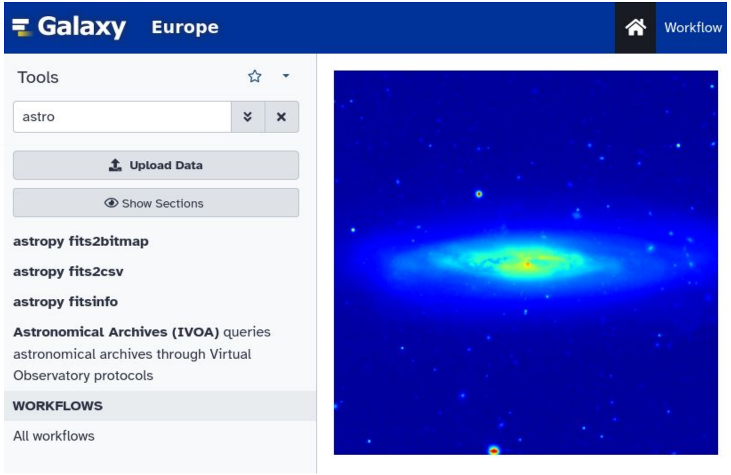
We took example of HDF5 to extend Galaxy with **FITS** format: **identification**, **parsing**, conversion to **tabular** to use **existing Galaxy tools**.

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





Most of the work by *Francois Morier-Genoud* https://galaxyproject.org/news/2023-06-20-esg-wp5-astronomy-fits/ Video

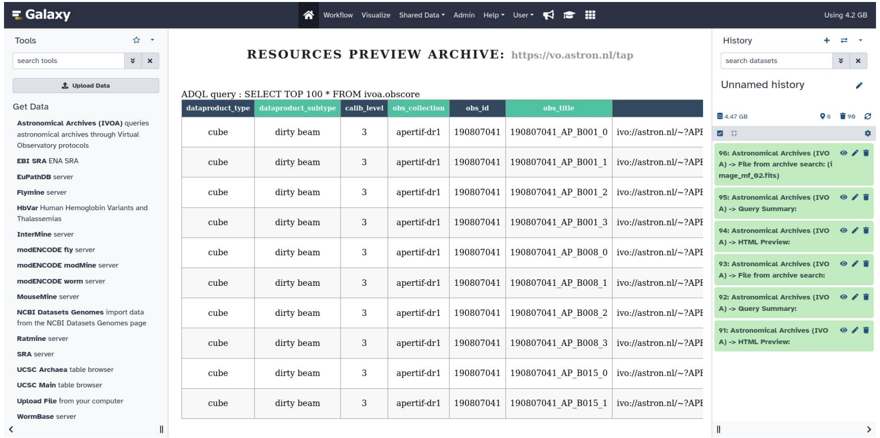
Adding Astro Data to Galaxy: IVOA archives

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 (**IVOA**) but also **ESA, NASA, ESO**.

We added first interface to query **IVOA 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



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.

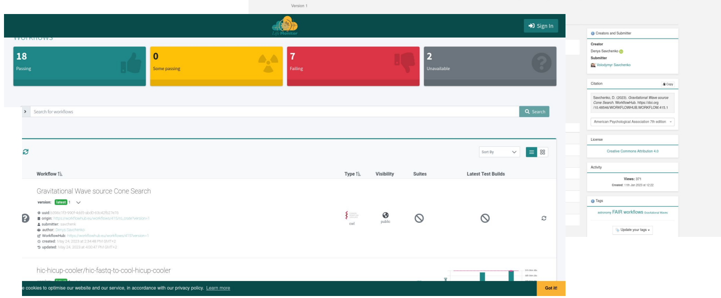
Another challenge is tracking **provenance of very large workflows**, often with many similar fragments.

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 poiting offset of less than three degrees from GX 5−1. This image can be obtained from the following URL ✓ (*ROCrate, 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

[LifeMonitor](#) [WorkflowHub](#)



```
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 \ch{which has an S/N of
    \VAR{search.grb.snr|round(1)}} in
    SPI-ACS. The blue shaded region corresponds to a range of one
    standard deviation of the background.}
\label{fig:acslc}
\end{figure}
```

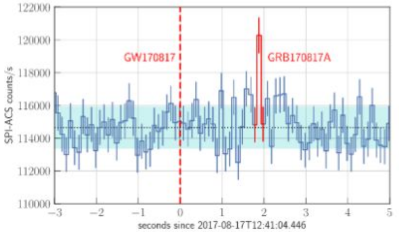
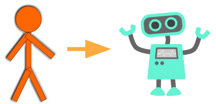


FIG. 2.— SPI-ACS light curve of GRB 170817A (100 ms time resolution), detected 2 seconds after GW170817. The red line highlights the 100 ms pulse, which has a S/N of 4.6 in SPI-ACS. The blue shaded region corresponds to a range of one standard deviation of the background.

Development space: help scientists make tools

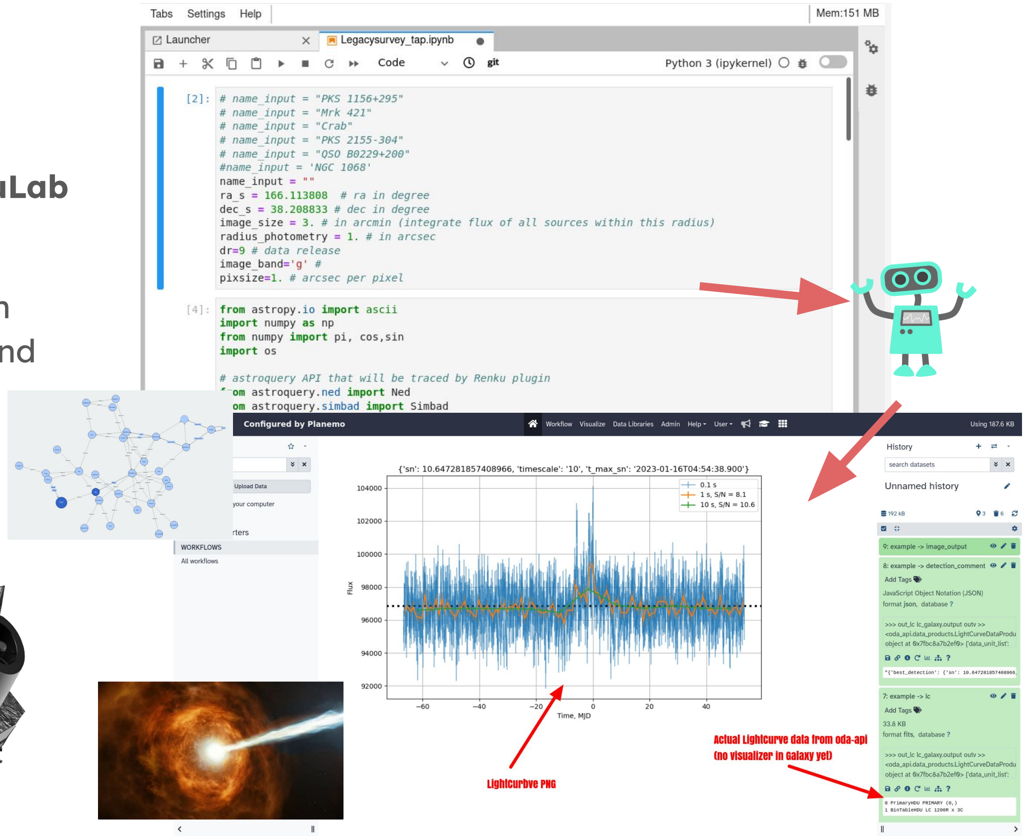


There are much **more scientists who can make a jupyter notebook than write organized code.**

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)



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:

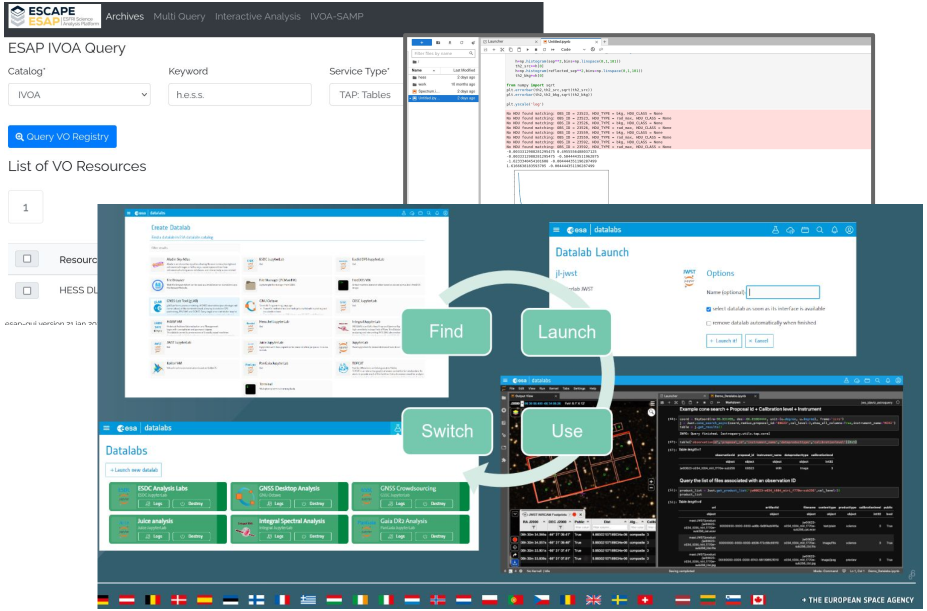
- Visualization plugins for data types
- Astro tool catalogs
- Data catalogs (IVOA, etc)
- Compute (e.g. ARC) and storage (Rucio) resources

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 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 End