MMODA multi-messenger online data analysis platform in the frame of the EuroScienceGateway project

Denys Savchenko[1,4]
Andrii Neronov[1,2], Cécile Cavet[1], Volodymyr Savchenko[2,3]

1. *Astroparticule et Cosmologie CNRS/UPCité*
2. *EPFL Laboratoire d'astrophysique*
3. *Observatoire Astronomique de lÚniversité de Genève*
4. *BITP & KAU, Ukraine*

Multi-Messenger Time Domain Astronomy

Exploding field!

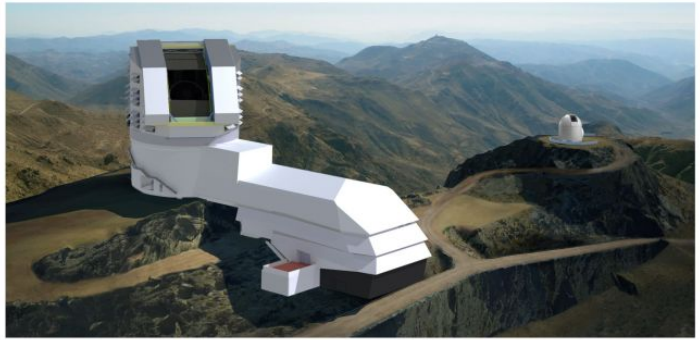
Last decade key **new observables** were discovered, and conventional telescopes dramatically upgraded to match.

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

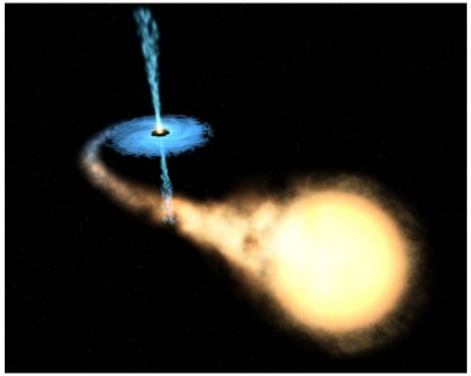
"Just" a star

Radio

Visible

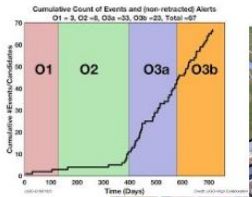


Star and black hole



Two neutron stars

Gravity



Multi-Messenger data analysis

Only combining data together, it is possible to see a complete picture of physical phenomena in astronomical sources.

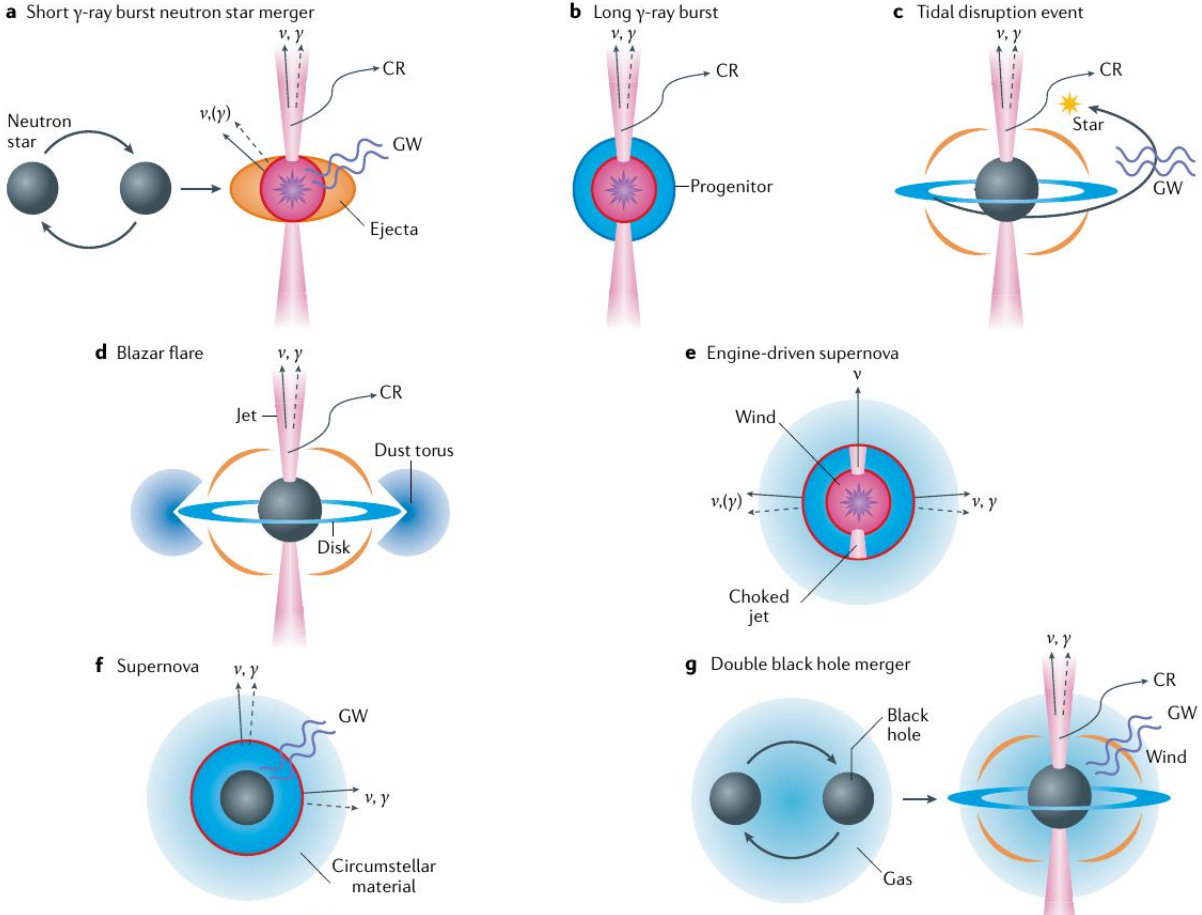
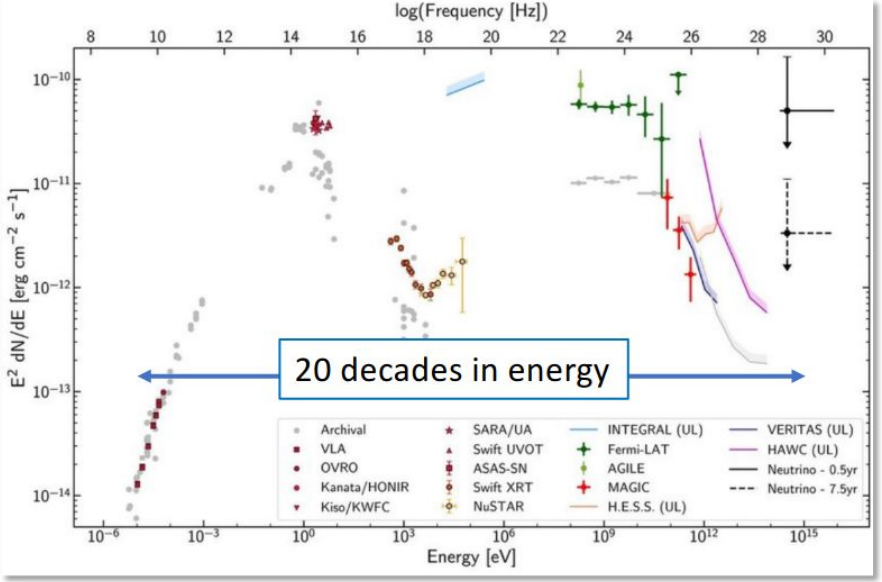
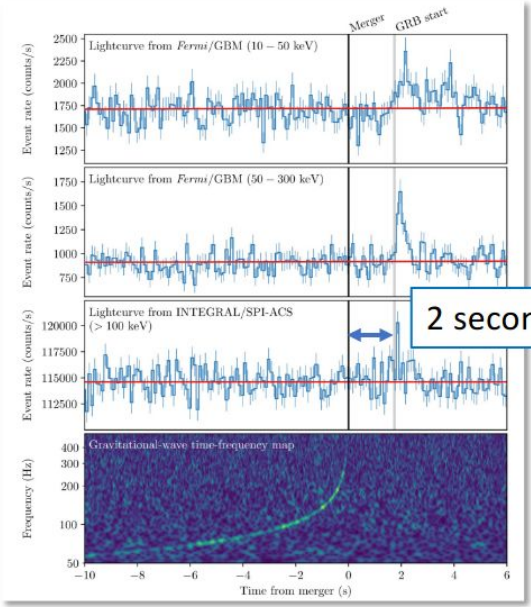


Fig: Meszaros et al. 2019

Multi-Messenger data analysis





Example: blazar TXS 0506+056 (a supermassive black hole).

A wealth of astronomical sources emit over very broad energy range. Understanding of emission mechanisms requires astronomical data collected with many different types of telescopes.

Individual astronomers cannot master data analysis techniques of all these telescopes at once. A system that helps (guides) them to obtain analysis-ready results for multiple types of astronomical instruments, would be useful.

Example: GW 170817 (a neutron star merger event).

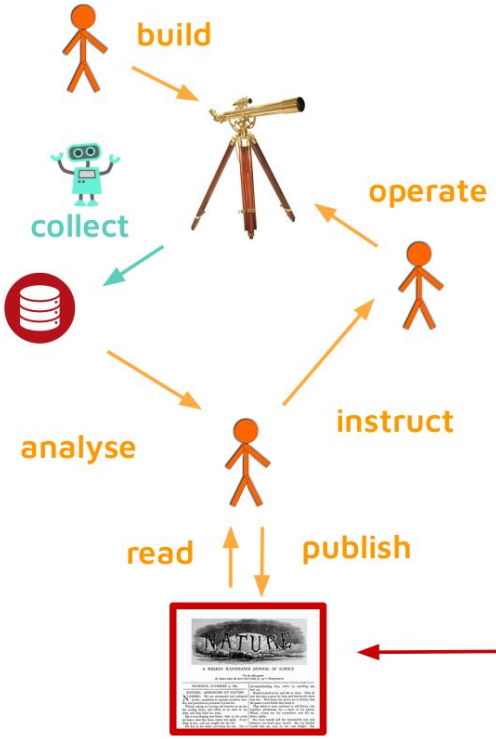
A wealth of astronomical sources appears on the sky for a short period of time (down to milli- and microseconds in the case of “fast radio bursts”). Understanding of emission mechanisms requires ”fast reaction”, to observe the source with multiple telescopes, while it is “in action”.

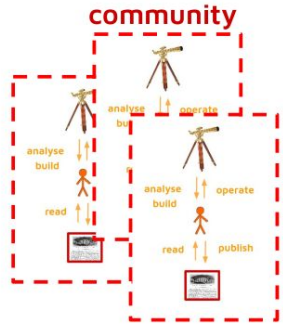
Individual astronomers cannot master all these telescopes at once. A system that helps coordinated observation campaigns and extracts data analysis results in automatic way would be useful.

When do we start

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**





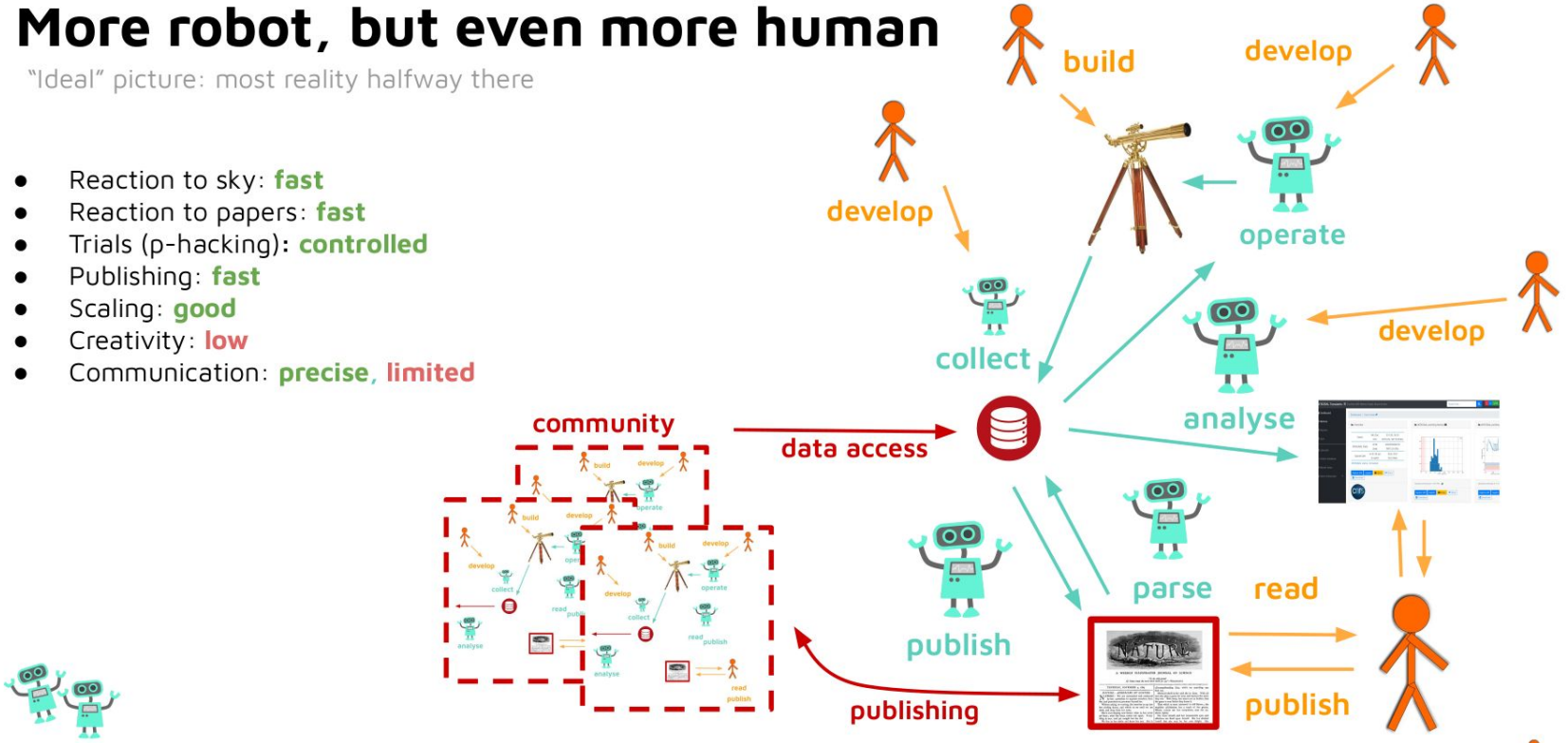
Human reaction and processing **is slow**, even if it's within even one person. But people are **smart**

Evolution

More robot, but even more human

"Ideal" picture: most reality halfway there

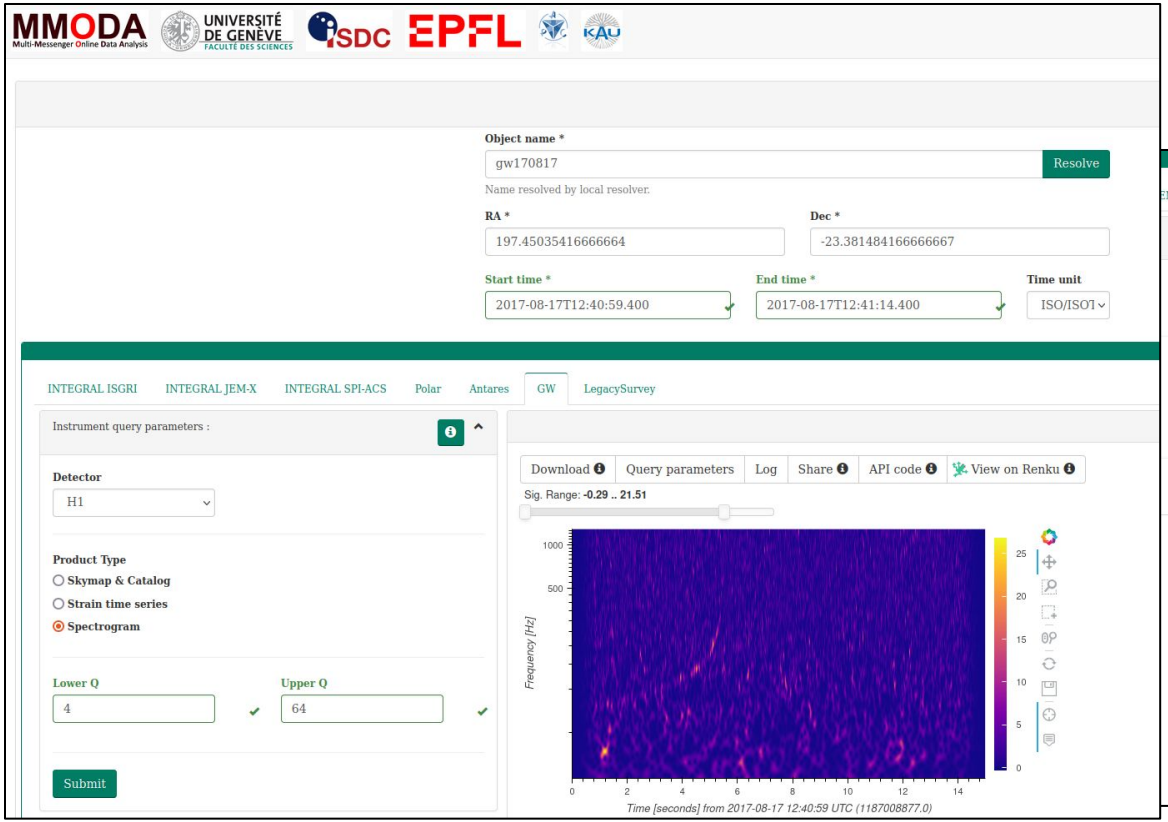
- Reaction to sky: **fast**
- Reaction to papers: **fast**
- Trials (p-hacking): **controlled**
- Publishing: **fast**
- Scaling: **good**
- Creativity: **low**
- Communication: **precise**, **limited**

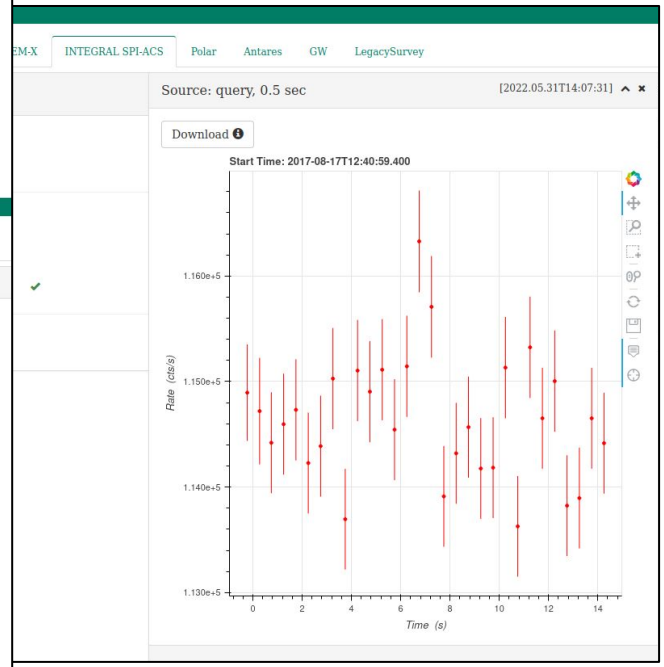


- **Making smart robots is hard**: always lacking **developers who are also research scientists.**
- If all is automated, **scientists have hard time seeing what's going on**, since **they do not speak robot**
- Robots are **fast**, but **lack creative reaction** in **new situations**.

MMODA: a tool for exploring, transforming MM data

https://www.astro.unige.ch/mmoda/





EUROPEAN OPEN SCIENCE CLOUD

https://marketplace.eosc-portal.eu/services/astronomical-online-data-analysis-astrooda

Software layer

https://github.com/oda-hub

-- **API** access using dedicated python library

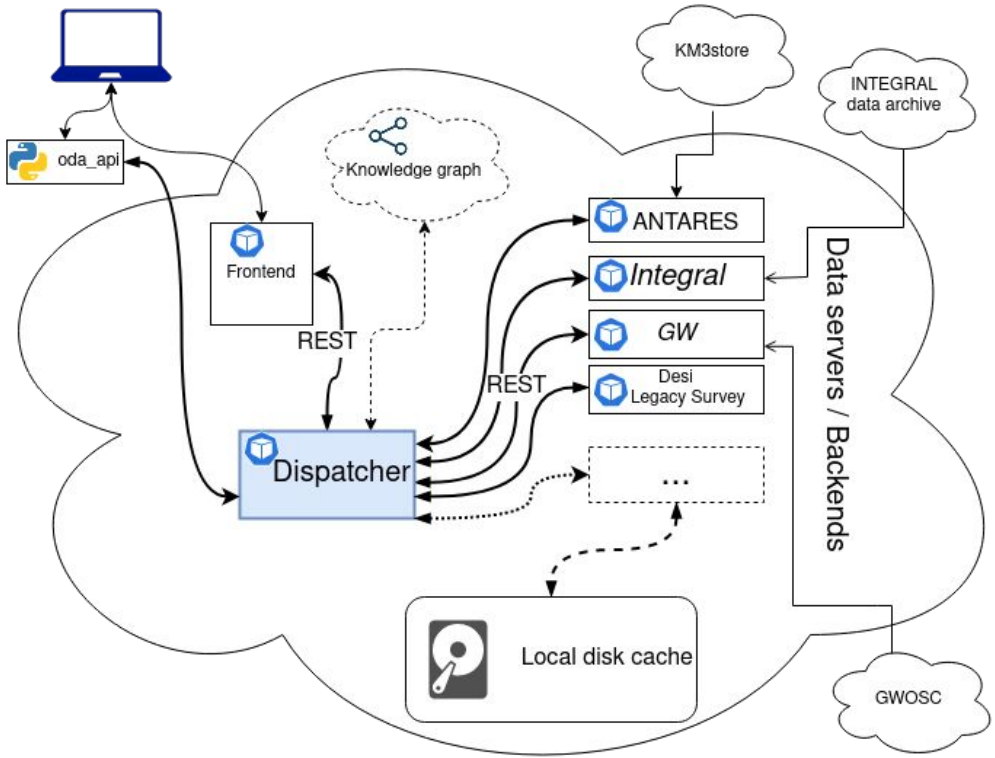
-- WEB-**frontend**

-- **dispatcher** coordinates data flow and job provisioning

-- data products are cached for later use

-- raw data from external services/archives

– provenance metadata in Knowledge Graph



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

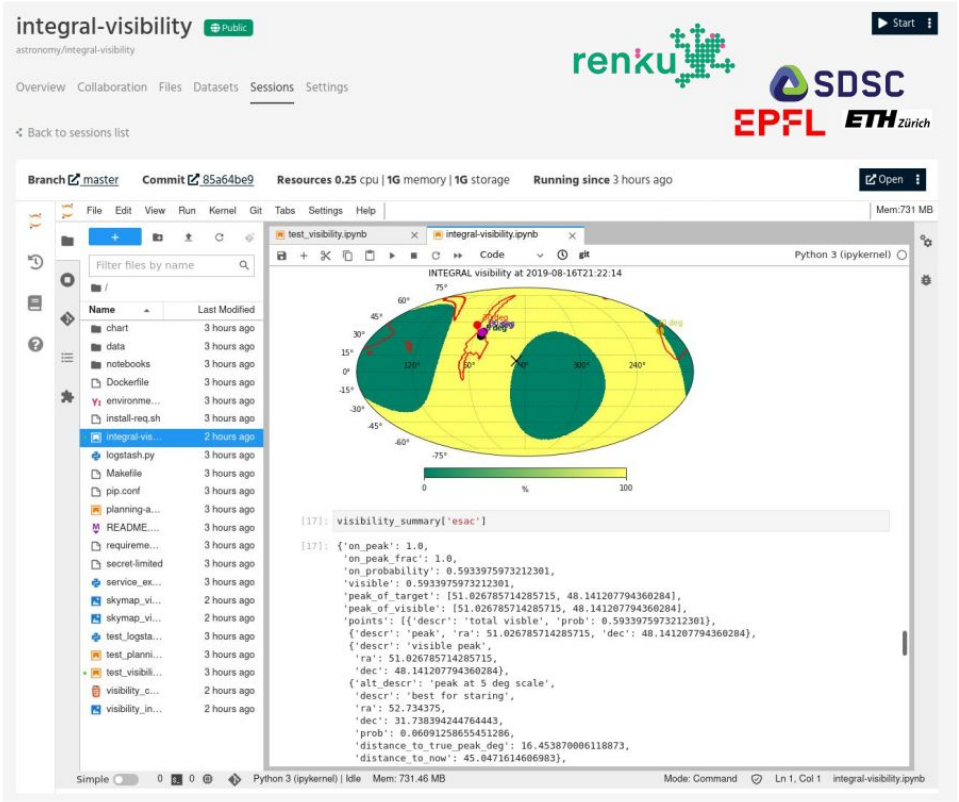
Development space: help scientists make robots

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

JupyterHub(s), Google-collab, ESA DataLabs, Renku

- **Continuous integration and testing**
- Supports in **publishing of data and code** (e.g. in zenodo)
- Support in **annotation** for scientists and robots reuse with ontology terms

This process creates a collection of notebooks and other workflows, but they are only really accessible interactively one-by-one

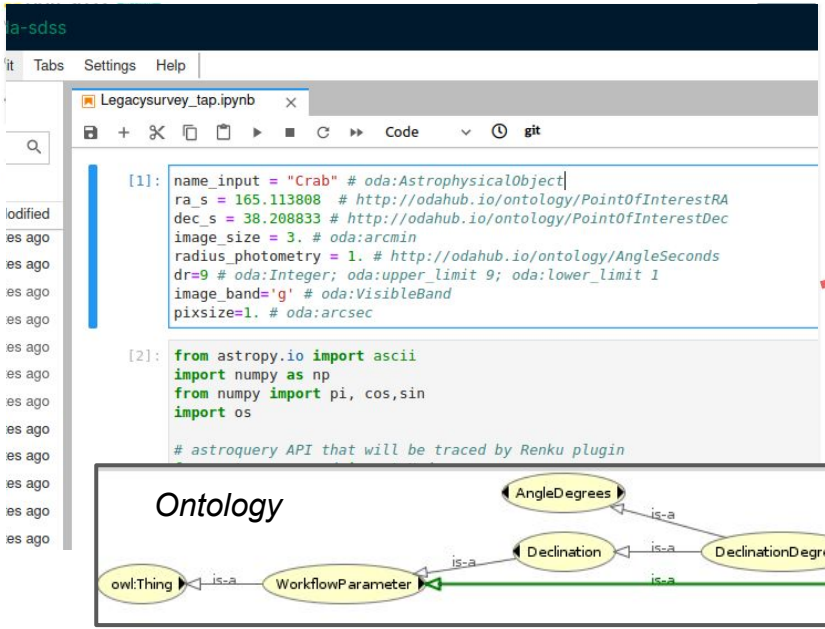


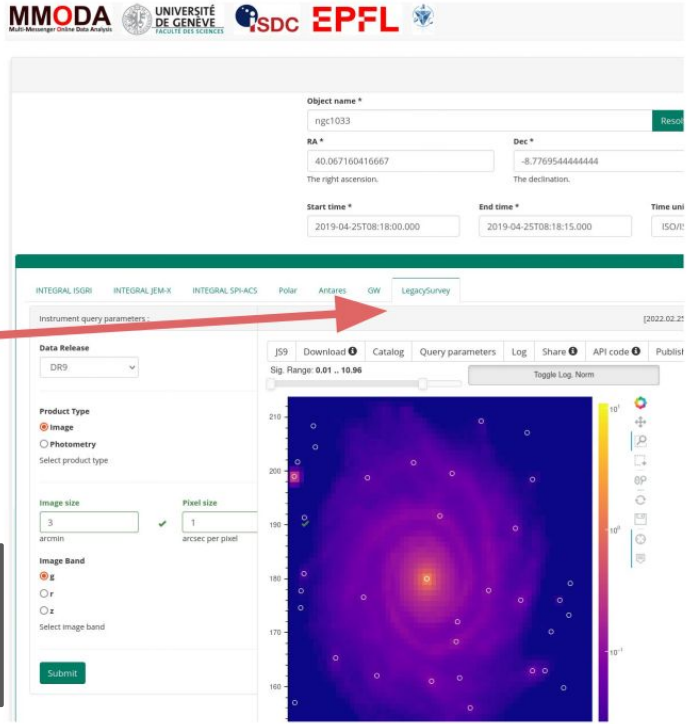
Making the developed workflow available as a web tool

jupyter may be easy, but sometimes we want just put parameters and click one button in **web interface**.

And even more so, we want to **leverage workflow as a service**, possibly calling from another workflow

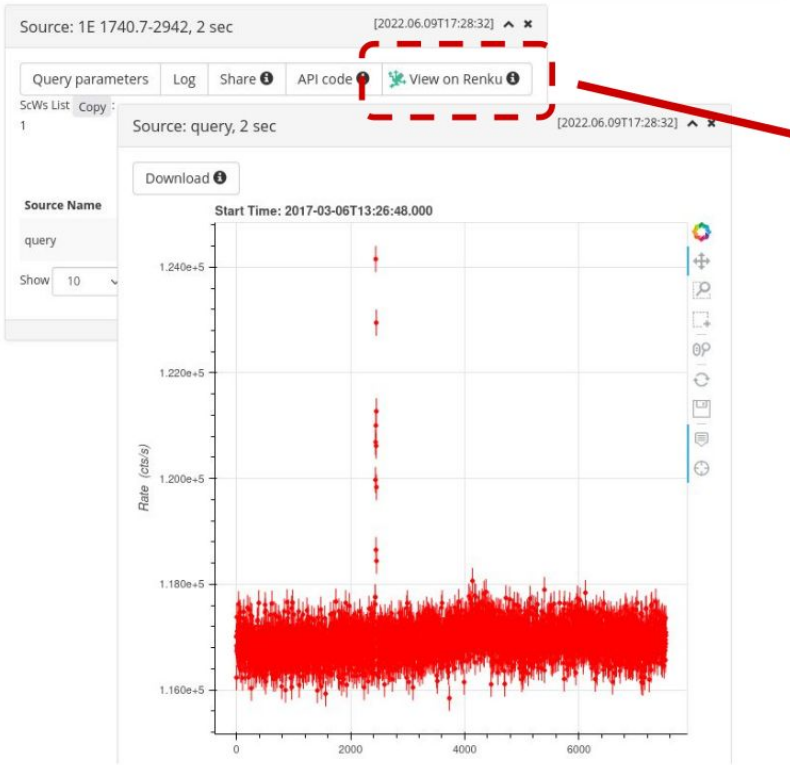
We are publishing the live tool, not just it's output

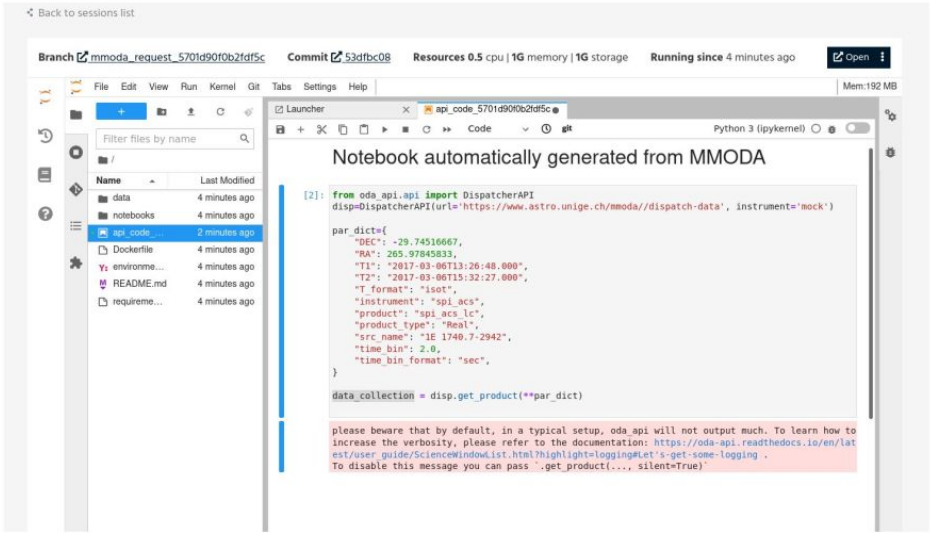




Helping to request MMODA services from Jupyter/Renku

Building new workflows by using results of the existing ones

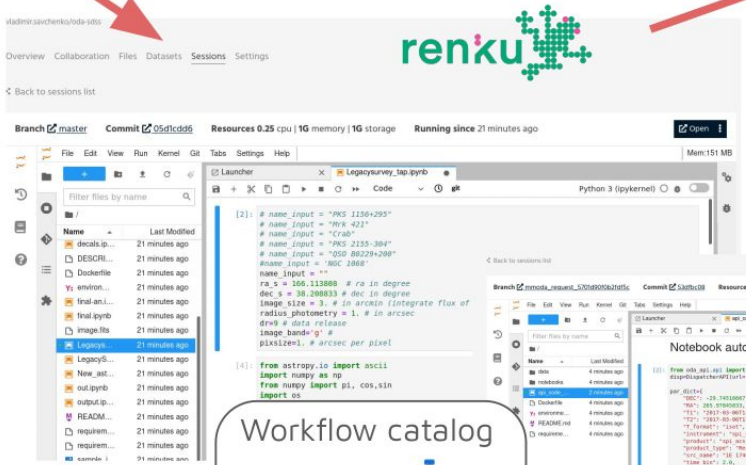


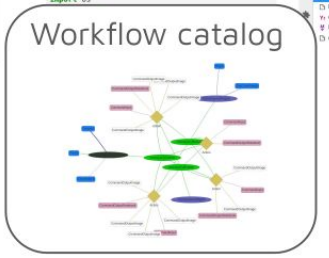


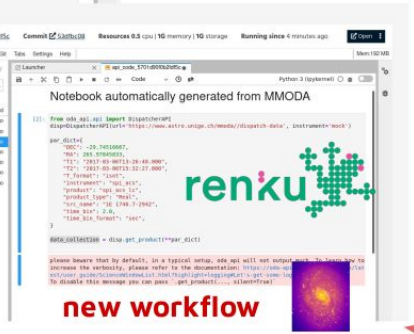
Feedback loop for crowd-sourcing workflow catalog

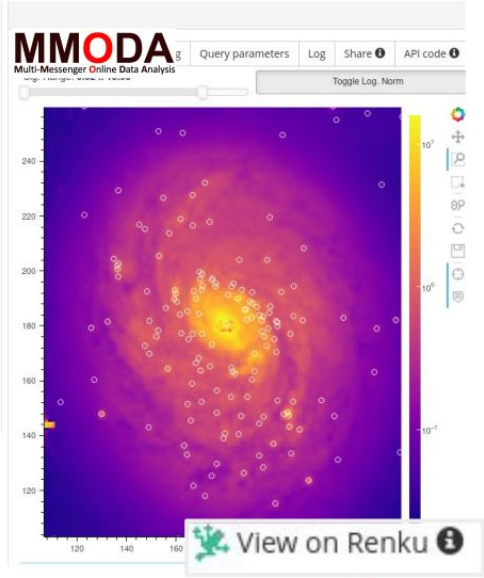
Scientist develops data reduction with deep scientific expertise

Publishing renku workflow as discoverable and executable asset









Automated workflow testing, benchmarking, reaction to space events, etc

Scientist creates new workflow leveraging the existing one

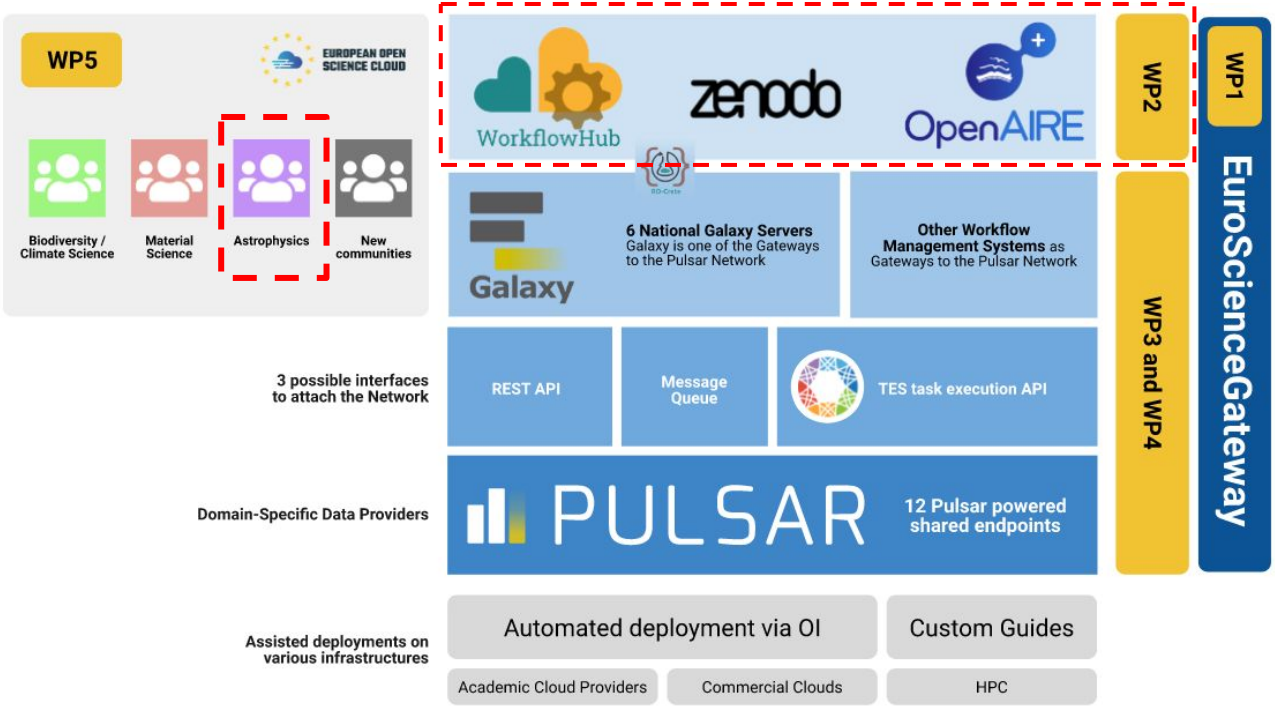
EuroScienceGateway

- 18 national and international institutions across 14 countries
- Lead by University of Freiburg (Germany)
- 3 years starting early 2023

EOSC project leveraging the European compute infrastructures for data-intensive research guided by FAIR principles

Key elements:

- "**Galaxy**" Web-based Science **Platform**
- Sustainable **Compute and Storage** network
- **FAIR** data and workflows: **publishing** and **preservation**
- Expanding **communities**



Galaxy data analysis platform

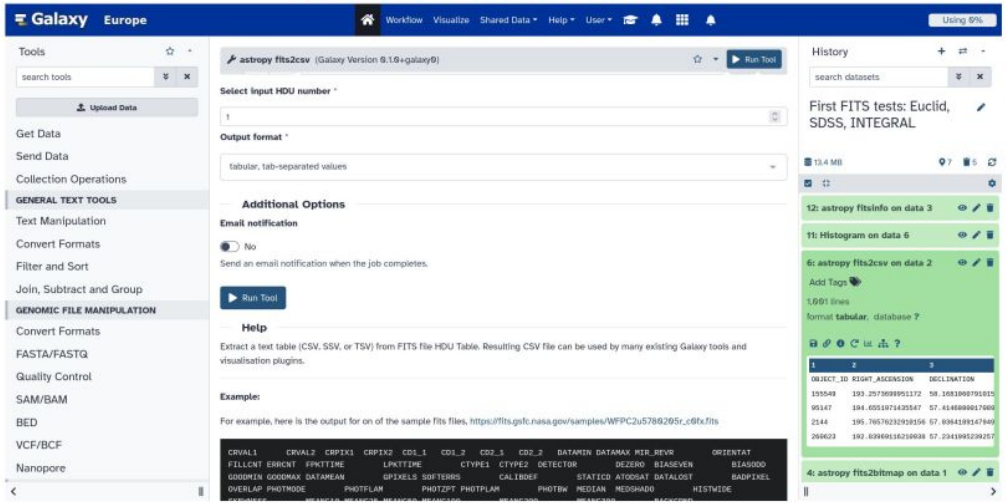
Grew out of bioinformatics needs, but reached **broad user community**: Life Sciences, Materials Science, Climate/Earth etc.

170 registered instances, dedicated well-developed training network

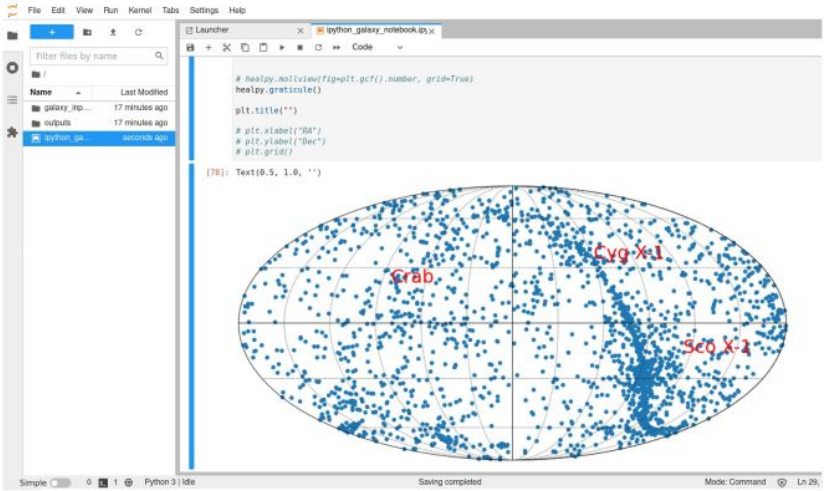
Very **modular and customizable**, data formats, visualization modules, job submission modules.

Explored **wide variety of design patterns** fit for different purposes

Unified User Interface



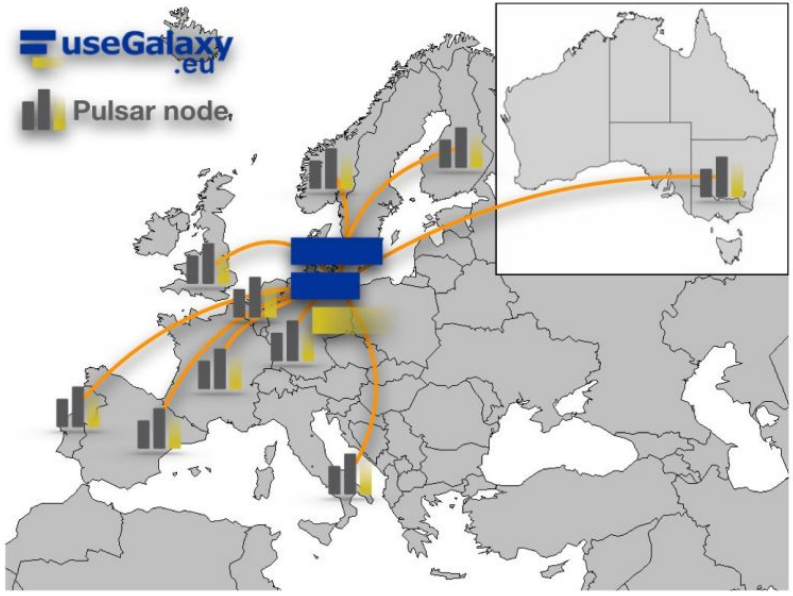
Interactive tools

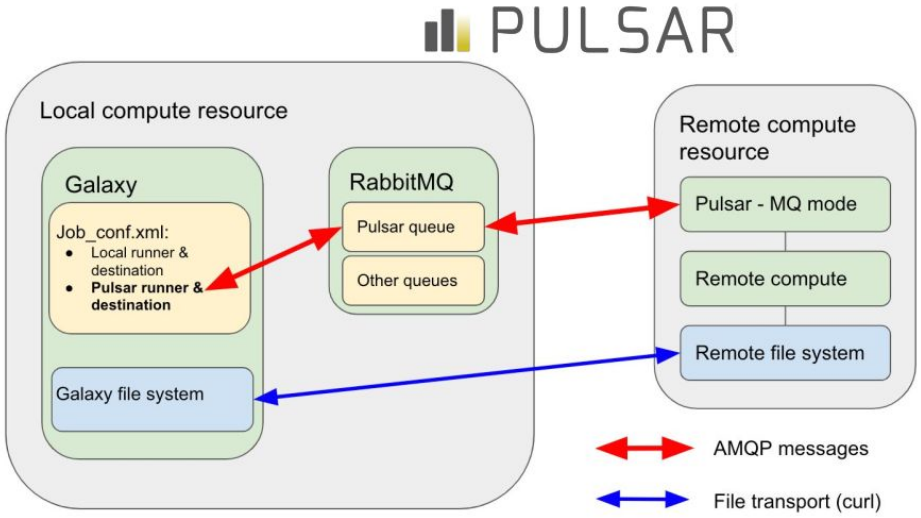


Pulsar distributed compute network

Pulsar services connected in a network enabling large computing network between **European supercomputing centers**

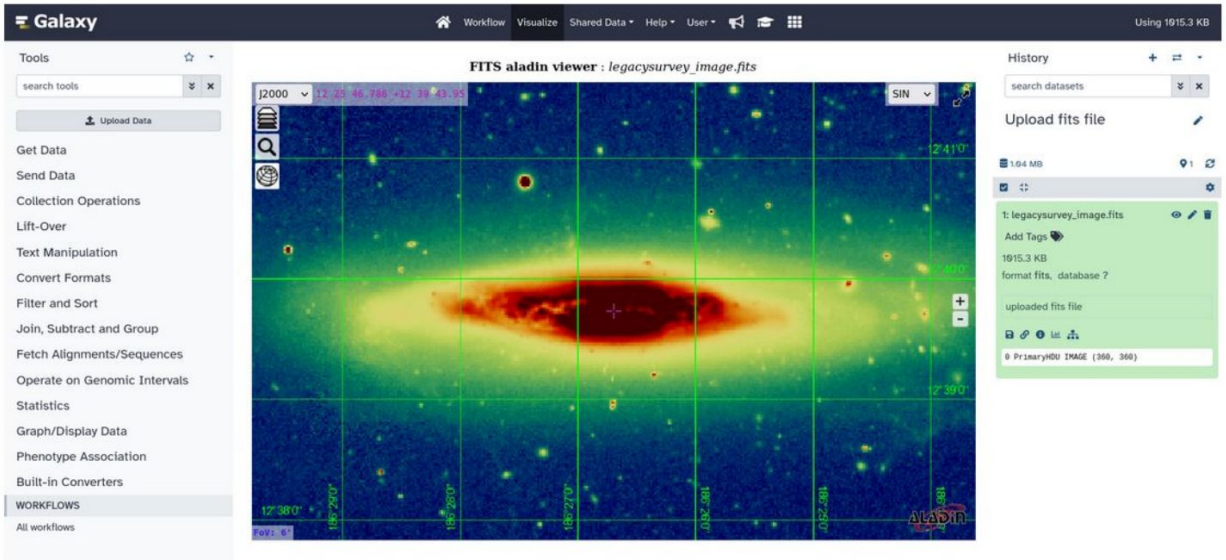
Bring Your Own Storage/Compute allows users to add their own resources to galaxy resource pool, ensuring **sustainable capacity**.





Galaxy Astronomy: FITS support, preview with AladinLite

Many common formats already supported. We extended Galaxy with **FITS** format: identification, parsing (astropy), and visualization (AladinLite).



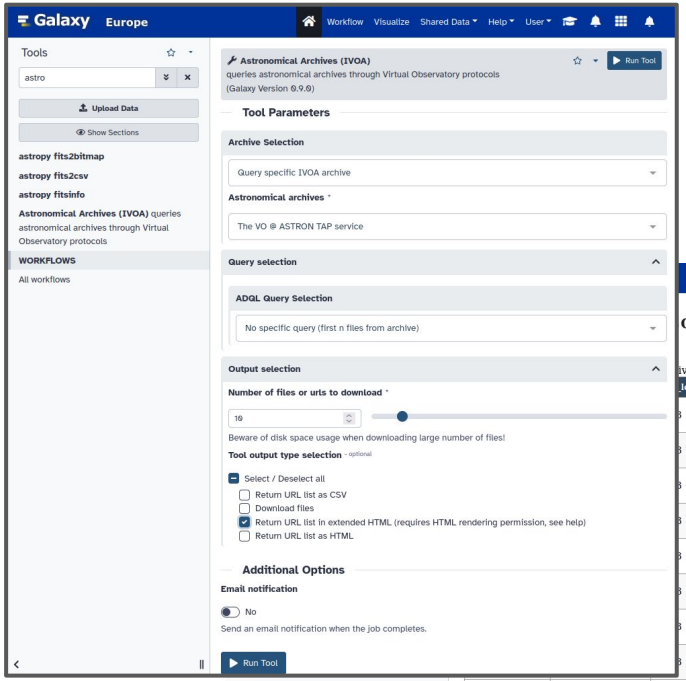
Most of the work by François Morier-Genoud (EPFL)

https://galaxyproject.org/news/2023-06-20-esg-wp5-astronomy-fits/

Adding IVOA archives

Galaxy supports S3, webdav, pyfilesystem, etc.

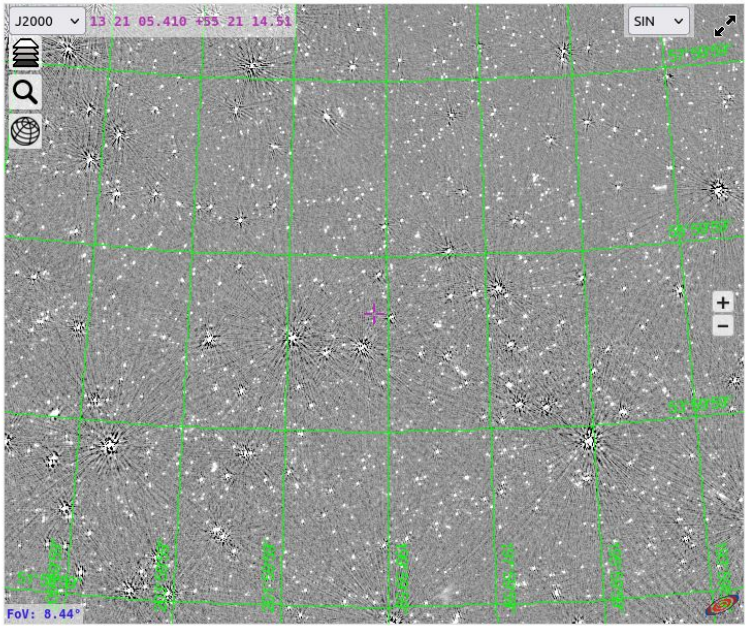
We added first interface to query **IVOA TAP archives** from Galaxy (demo video).

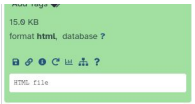


OURCES PREVIEW ARCHIVE: https://vo.astron

ivoa.obscore

level	obs_collection	obs_id	obs_title	
3	apertif-dr1	190807041	190807041_AP_B001_0	ivo://astron.nl/~
3	apertif-dr1	190807041	190807041_AP_B001_1	ivo://astron.nl/~
3	apertif-dr1	190807041	190807041_AP_B001_2	ivo://astron.nl/~
3	apertif-dr1	190807041	190807041_AP_B001_3	ivo://astron.nl/~
3	apertif-dr1	190807041	190807041_AP_B008_0	ivo://astron.nl/~
3	apertif-dr1	190807041	190807041_AP_B008_1	ivo://astron.nl/~
3	apertif-dr1	190807041	190807041_AP_B008_2	ivo://astron.nl/~
3	apertif-dr1	190807041	190807041_AP_B008_3	ivo://astron.nl/~?APERTIF_DR1/190807041_AP_B008/HI_image
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B015_0	ivo://astron.nl/~?APERTIF_DR1/190807041_AP_B015/HI_image
cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B015_1	ivo://astron.nl/~?APERTIF_DR1/190807041_AP_B015/HI_image





Most of the work by François Morier-Genoud (EPFL)

Bringing our workflow catalogue into Galaxy platform

We are developing the nb2galaxy tool

The bot will convert our crowdsourced workflows in the form of python **notebooks to Galaxy tools** and add to the toolshed just like deploying them as a services in MMODA

detectgrb.ipynb 98.54 KiB Edit

In [1]:
```
T1 = "2023-01-16T04:53:33.9" # http://odahub.io/ontology#StartTime
T2 = "2023-01-16T04:55:33.9" # http://odahub.io/ontology#EndTime
detection_time_scales = "1,10"
lc_time_scale = 0.1 # https://odahub.io/ontology#TimeIntervalSeconds
background_age = 10 # oda:TimeIntervalSeconds
min_sn = 5 # https://odahub.io/ontology#SignalToNoiseRatio
```

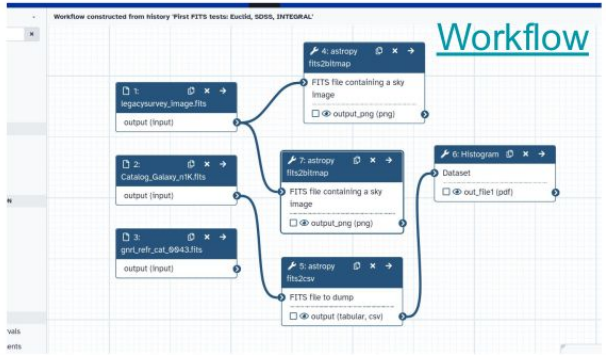
In [2]:
```
import numpy as np
from astropy.time import Time
from oda_api.data_products import LightCurveDataProduct, BinaryData, PictureProduct
from matplotlib import pylab as plt
```

In [3]:
```
from oda_api.api import DispatcherAPI
disp=DispatcherAPI(url='https://www.astro.unige.ch/mmoda//dispatch-data', instrument='mock')

par_dict={
    "T1": T1,
    "T2": T2,
    "T_format": "isot",
    "instrument": "spi_acs",
    "product": "spi_acs_lc",
    "product_type": "Real",
    "time_bin": lc_time_scale,
    "time_bin_format": "sec",
}

data_collection = disp.get_product(**par_dict)

lc = data_collection.spi_acs_lc_0_query.data_unit[1].data
```

Out [3]:

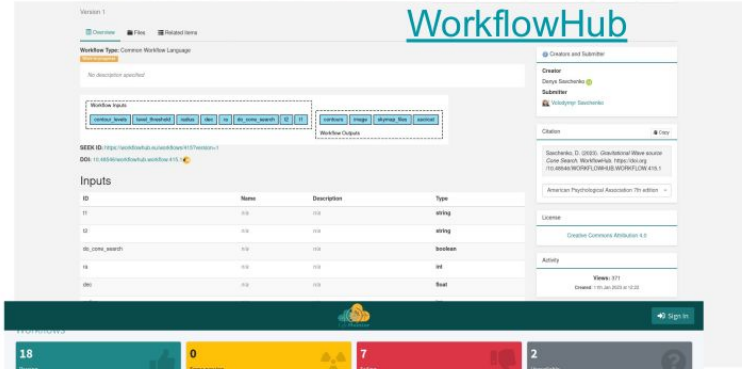
FAIR Workflows Catalogs, Publishing

Workflow is **more** than **software,** it has **machine-readable instructions to execute**



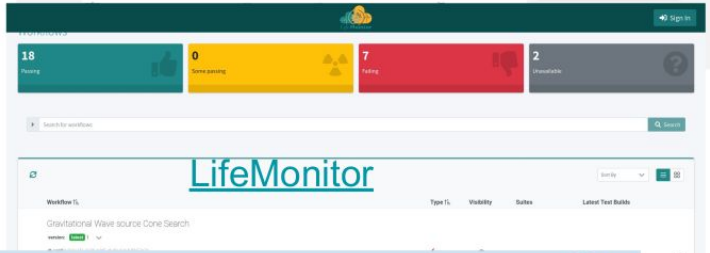
RO-Crate represents workflow in publishable form, with **semantic** annotations

Stored in discoverable **workflow catalog:** WorkflowHub



We are **connecting tool catalogs** by ingesting tools developed in SDSC **RenkuLab** and with **MMODA** Multi-Messenger Online Data Analysis (**AstroORDAS)**

Recuperability: need make sure workflow is still alive when it has been published a while ago.



Workflows are then be **embedded** into journal **publications** and published with **DOI** in archives and registries.

Provenance-first *INTEGRAL/MMODA paper converted into an example.*

Summary

- Modern astronomy (especially multi-messenger) is all about **rapidly growing data** and **reducing the time** of the analysis
- Need **intelligent automation** to react fast and ensure **reuse and reproducibility**
- We establish an **ecosystem** centered around MMODA platform, which allows to **crowd-source FAIR workflow** creation
- It's always beneficial to leverage **synergies with other projects**
- In EuroScienceGateway we **integrate existing solutions and workflow** catalogs with **Galaxy**
- Galaxy is a very **flexible, well-developed science platform** with exceptional experience, and it is highly beneficial to learn from it
- EuroScienceGateway project will help to see the **future potential of Galaxy platform for astronomy**