

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

Denys Savchenko^{1,4}

Andrii Neronov^{1,2}, Cécile Cavet¹, Volodymyr Savchenko^{2,3}

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

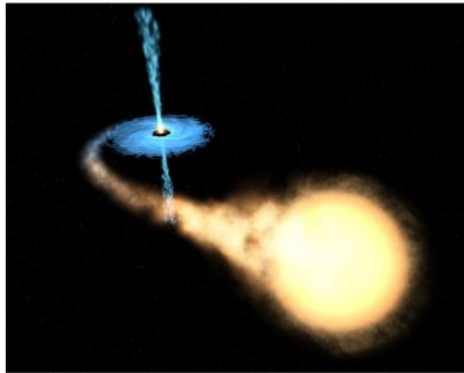
Multi-Messenger Time Domain Astronomy

Exploding field!

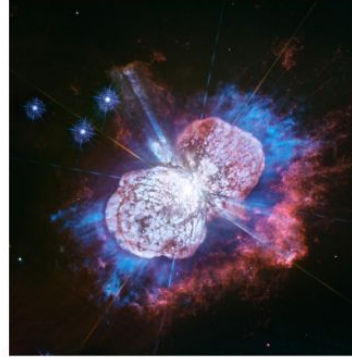
Last decade key **new observables** were discovered, and conventional telescopes dramatically upgraded to match.

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

Star and black hole



"Just" a star



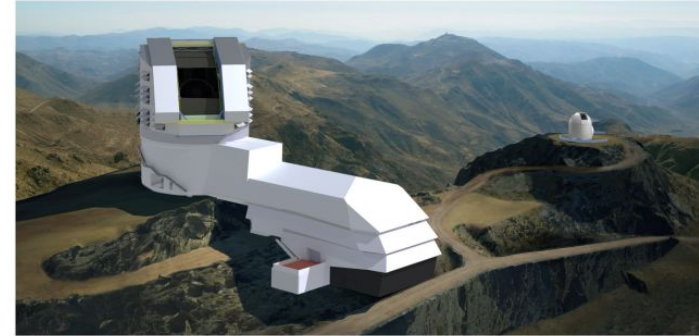
Two neutron stars



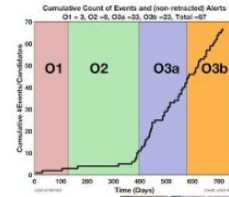
Radio



Visible



Gravity



Multi-Messenger data analysis

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

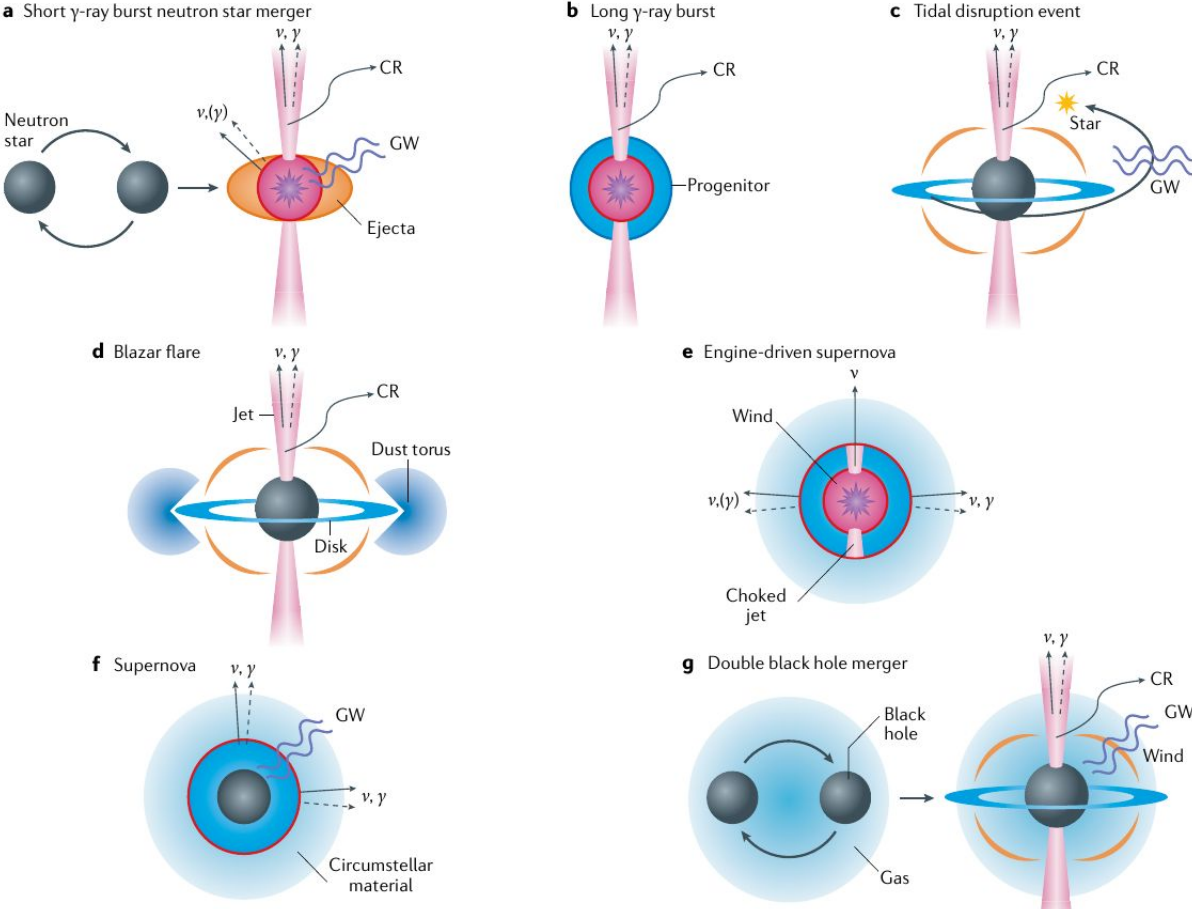
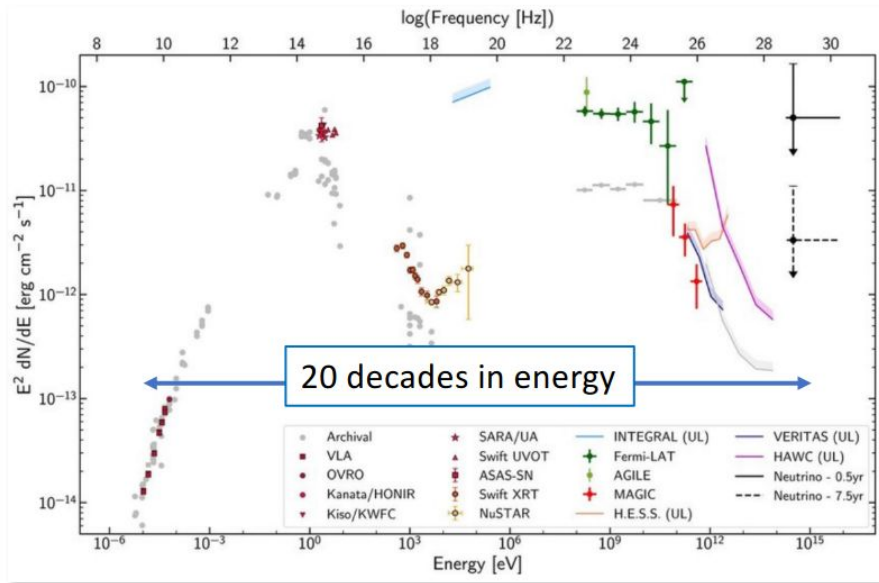


Fig: Meszaros et al. 2019

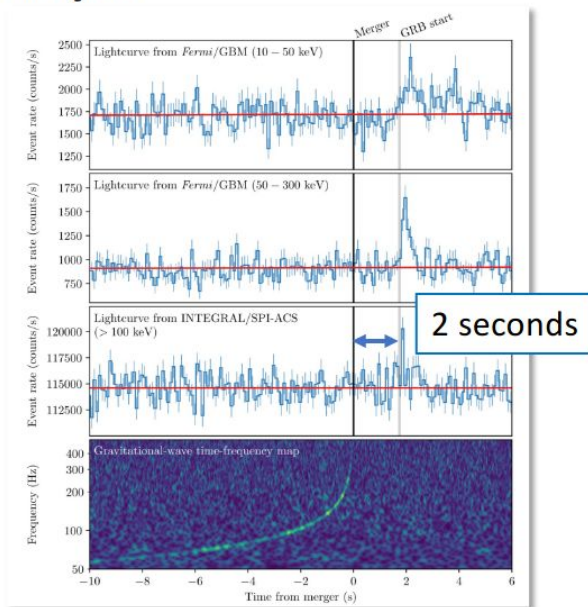
Multi-Messenger data analysis



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).

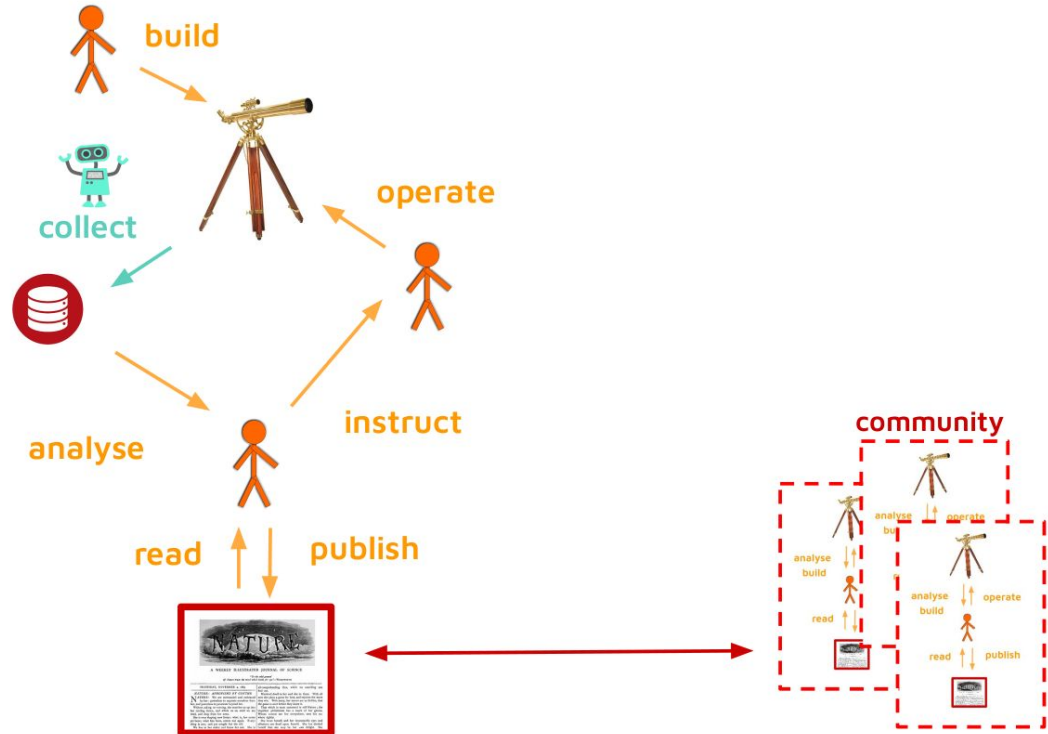
A wealth of astronomical sources appears on the sky for a short period of time (down to milli- and microseconds in the case of “fast radio bursts”). Understanding of emission mechanisms requires “fast reaction”, to observe the source with multiple telescopes, while it is “in action”.

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**



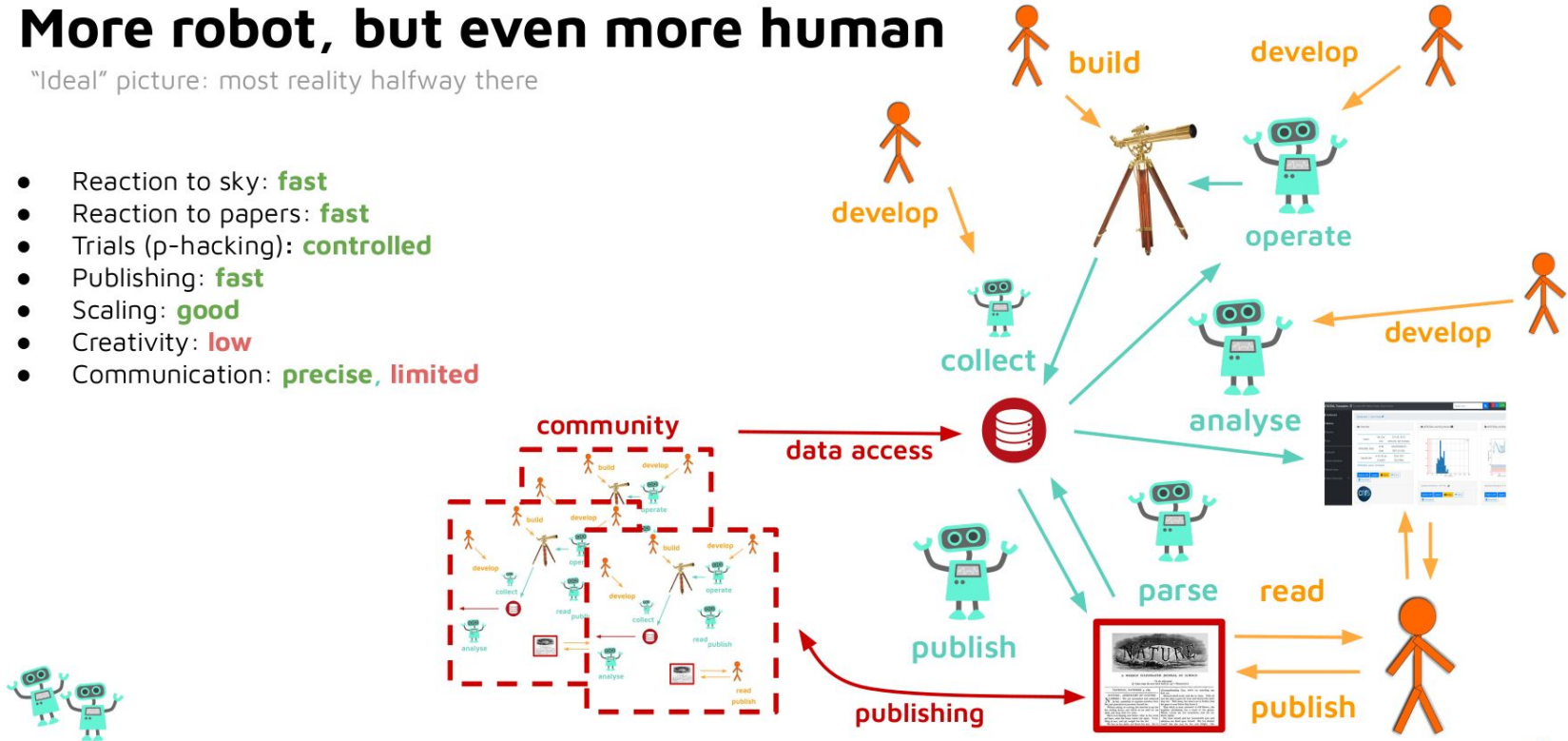
Human reaction and processing **is slow**, even if it's within even one person. But people are **smart**

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**.

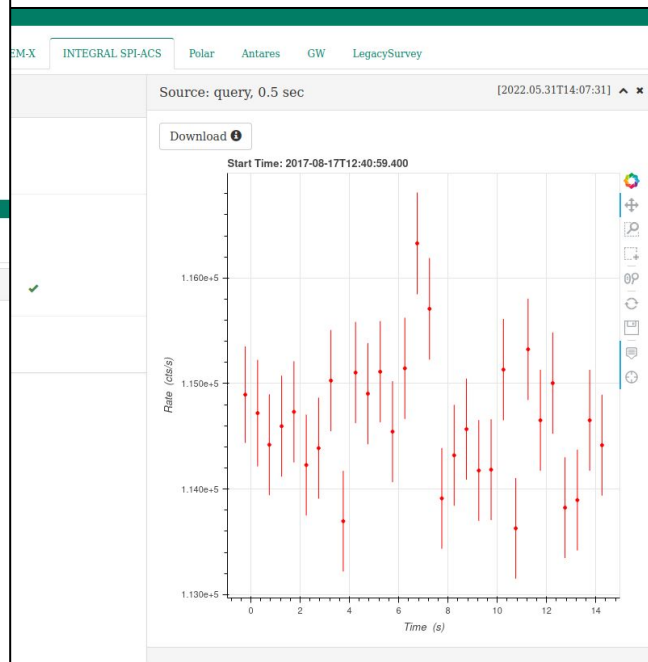
MMODA: a tool for exploring, transforming MM data

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

The screenshot shows the MMODA web interface. At the top, there are logos for MMODA (Multi-Messenger Online Data Analysis), UNIVERSITÉ DE GENÈVE FACULTÉ DES SCIENCES, ISDC, EPFL, and KAU. The main search area includes fields for Object name (*), RA (*), Dec (*), Start time (*), End time (*), and Time unit. The object name is 'gw170817'. Below the search area, there are tabs for different instruments: INTEGRAL ISGRI, INTEGRAL JEM-X, INTEGRAL SPI-ACS, Polar, Antares, GW, and LegacySurvey. The 'GW' tab is selected. On the left, there are 'Instrument query parameters' including a 'Detector' dropdown (set to 'H1'), 'Product Type' options (Skymap & Catalog, Strain time series, Spectrogram), and 'Lower Q' and 'Upper Q' input fields (set to 4 and 64). A 'Submit' button is at the bottom left. The main plot area shows a spectrogram with 'Frequency [Hz]' on the y-axis (0 to 1000) and 'Time [seconds] from 2017-08-17 12:40:59 UTC (1187008877.0)' on the x-axis (0 to 14). A color bar on the right indicates intensity from 0 to 25. Above the plot are buttons for 'Download', 'Query parameters', 'Log', 'Share', 'API code', and 'View on Renku'. A 'Sig. Range: -0.29 .. 21.51' slider is also present.

MMODA

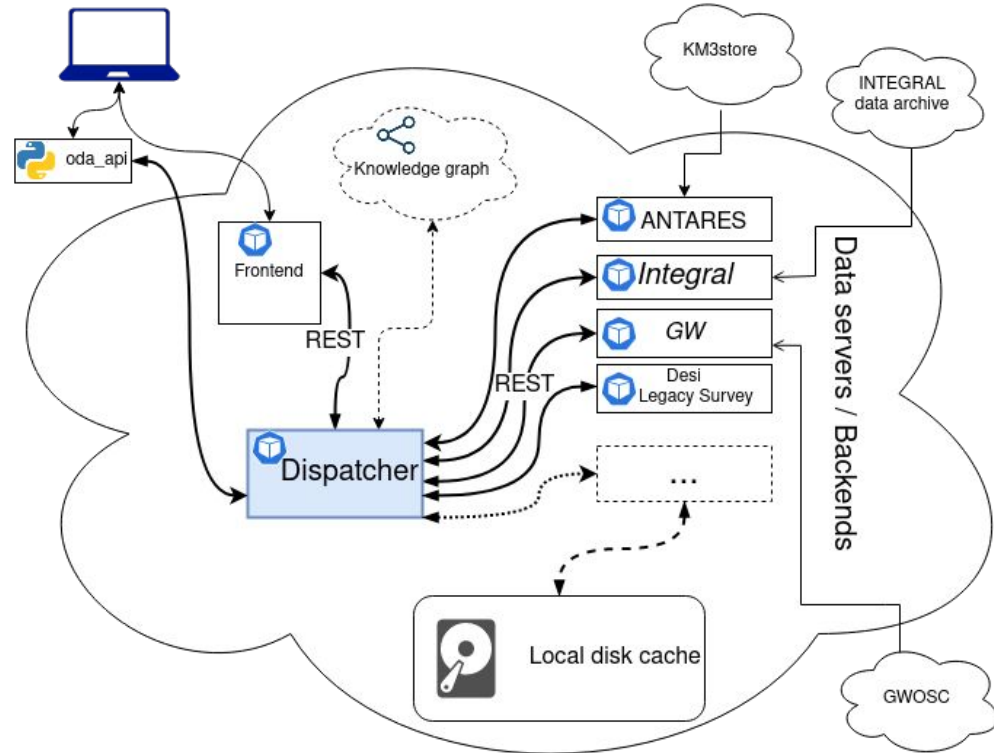
Multi-Messenger Online Data Analysis



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

integral-visibility Public
astronomy/integral-visibility

Overview Collaboration Files Datasets Sessions Settings

← Back to sessions list

Branch **master** Commit **95a64be9** Resources **0.25** cpu | **1G** memory | **1G** storage Running since 3 hours ago

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

Name	Last Modified
chart	3 hours ago
data	3 hours ago
notebooks	3 hours ago
Dockerfile	3 hours ago
environme...	3 hours ago
install-req.sh	3 hours ago
integral-vis...	2 hours ago
logstash.py	3 hours ago
Makefile	3 hours ago
pip.conf	3 hours ago
planning-...	3 hours ago
README...	3 hours ago
requireme...	3 hours ago
secret-limited	3 hours ago
service_ext...	3 hours ago
skymap_vi...	2 hours ago
skymap_vi...	2 hours ago
test_logsta...	3 hours ago
test_planni...	3 hours ago
test_visibili...	3 hours ago
visibility_c...	2 hours ago
visibility_in...	2 hours ago

INTEGRAL visibility at 2019-08-16T21:22:14

visibility_summary['esac']

```
[17]: {'on_peak': 1.0,
'on_peak_frac': 1.0,
'on_probability': 0.5933975973212301,
'visible': 0.5933975973212301,
'peak_of_target': [51.026785714285715, 48.141207794360284],
'peak_of_visible': [51.026785714285715, 48.141207794360284],
'points': [{'descr': 'total visible', 'prob': 0.5933975973212301},
{'descr': 'peak', 'ra': 51.026785714285715, 'dec': 48.141207794360284},
{'descr': 'visible peak',
'ra': 51.026785714285715,
'dec': 48.141207794360284},
{'descr': 'peak at 5 deg scale',
'ra': 52.734375,
'dec': 31.738384244764443},
{'descr': 'best for staring',
'ra': 52.734375,
'dec': 31.738384244764443},
{'prob': 0.06091258655451286,
'distance_to_true_peak_deg': 16.453870006118873,
'distance_to_now': 45.0471614666983}]
```

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



Object name *
ngc1033

RA *
40.067160416667

Dec *
-8.77695444444444

The right ascension. The declination.

Start time *
2019-04-25T08:18:00.000

End time *
2019-04-25T08:18:15.000

Time unit
ISO7

INTEGRAL ISGR INTEGRAL JEM-X INTEGRAL SPI-ACS Polar ARCADE GW LegacySurvey

Instrument query parameters: [2022.02.22]

Data Release
DR9

Product Type
 Image
 Photometry
Select product type

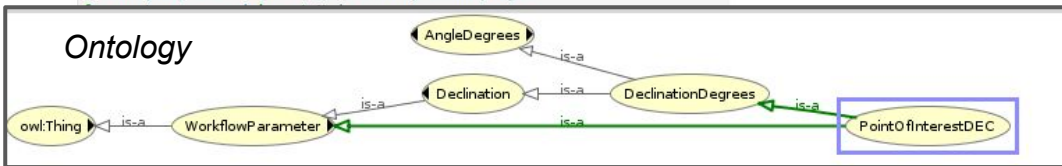
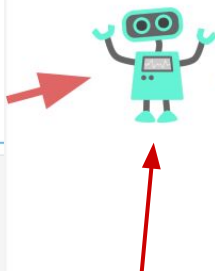
Image size
3 arcmin

Pixel size
1 arcsec per pixel

Image Band
 g
 r
 z
Select image band

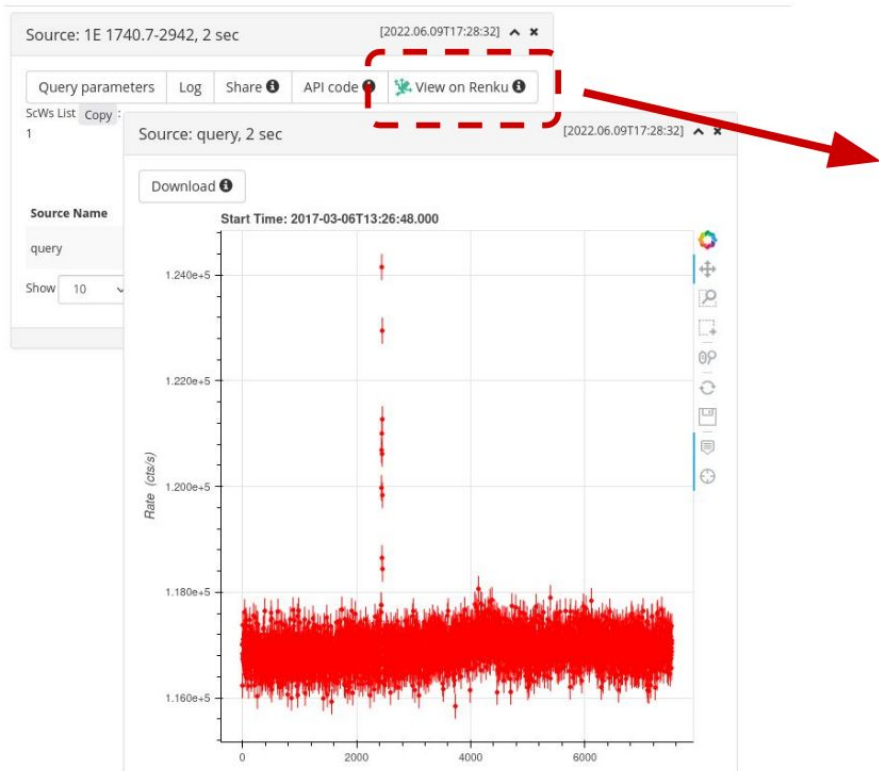
Submit

J59 Download Catalog Query parameters Log Share API code Publish
Sig. Range: 0.01 .. 10.96
Toggle Log Norm



Helping to request MMODA services from Jupyter/Renku

Building new workflows by using results of the existing ones



Branch `mmoda_request_5701d90fb2df5c` Commit `53dfbc08` Resources 0.5 cpu | 1G memory | 1G storage Running since 4 minutes ago

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

par_dict={
    "DEC": -29.74516667,
    "RA": 265.97845833,
    "TI": "2017-03-06T13:26:48.000",
    "T2": "2017-03-06T15:32:27.000",
    "I_format": "isot",
    "instrument": "spi_acs",
    "product": "spi_acs_lc",
    "product_type": "Real",
    "src_name": "1E 1740.7-2942",
    "time_bin": 2.0,
    "time_bin_format": "sec",
}

data_collection = disp.get_product(**par_dict)
```

please beware that by default, in a typical setup, oda api will not output much. To learn how to increase the verbosity, please refer to the documentation: https://oda-api.readthedocs.io/en/latest/user_guide/ScienceWindowList.html#highlight=logging&let's-get-some-logging . To disable this message you can pass `get_product(..., silent=True)`

Feedback loop for crowd-sourcing workflow catalog

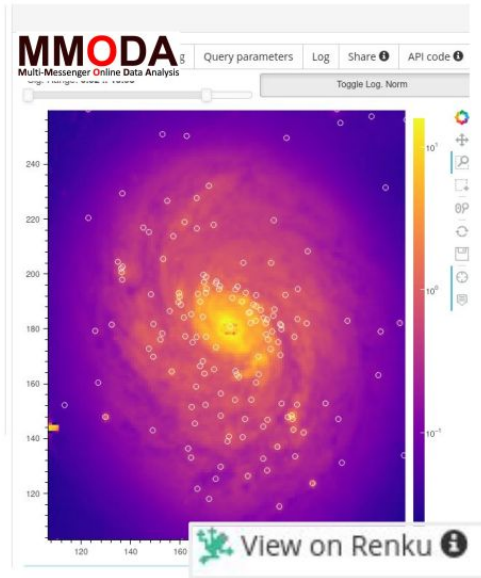
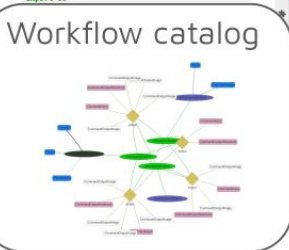


Scientist develops data reduction with deep scientific expertise

Publishing renku workflow as discoverable and executable asset



The screenshot shows a Renku workflow editor interface. At the top, there's a navigation bar with 'Overview', 'Collaboration', 'Files', 'Datasets', 'Sessions', and 'Settings'. Below that, a file browser shows a list of files like 'decals.ip...', 'Dockerfile', 'envron.', 'final.ipynb', 'image.fits', 'Legacyis...', 'Legacyis...', 'New_ast...', 'out.ipynb', 'outofp...', 'README...', 'requirem...', 'requirem...', and 'example-1'. The main area displays a code editor with Python code for a workflow. The code includes comments and function definitions for data reduction, such as `from astropy.io import ascii`, `import numpy as np`, and `from numpy import pi, cos, sin`. The code is part of a workflow named 'Legacyisurvey_step.ipynb'.



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



Scientist creates new workflow leveraging the existing one

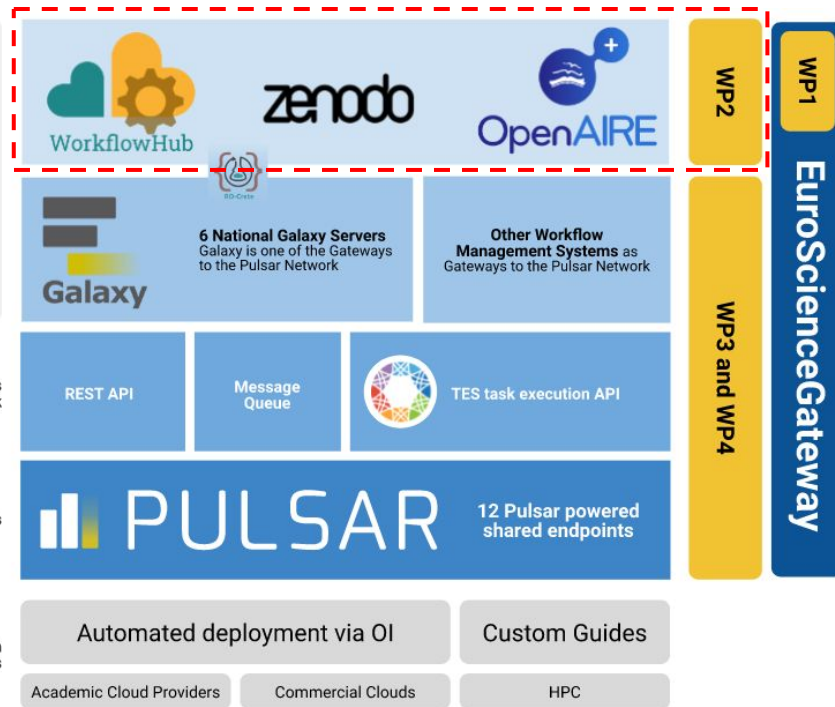
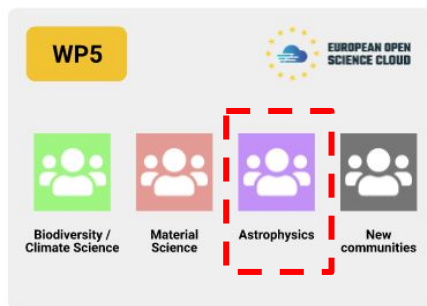
EuroScienceGateway

- 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

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy Europe', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'User', and 'Using 99%'. The main content area is titled 'astropy fits2csv (Galaxy Version 9.1.6+galaxy9)'. It features a 'Tools' sidebar on the left with categories like 'GENERAL TEXT TOOLS' and 'GENOMIC FILE MANIPULATION'. The central workspace shows a workflow with a 'Run Tool' button. Below the workflow, there is a 'History' panel listing previous jobs, including 'First FITS tests: Euclid, SDSS, INTEGRAL'. At the bottom, a table displays the output of the tool, with columns for 'OBJECT_ID', 'RA', and 'DECLINATION'.

OBJECT_ID	RA	DECLINATION
135049	193.25736989951372	58.36830887938102
95147	194.4001973423587	57.45488884674680
2144	195.7676202029588	57.0848181647440
268623	192.439686182326938	57.23419952982157

Interactive tools

The screenshot shows an interactive tool interface. The top navigation bar includes 'File', 'Edit', 'View', 'Run', 'Kernel', 'Tabs', 'Settings', and 'Help'. The main content area is titled 'Launcher' and shows a code editor with Python code for plotting galaxy positions. The code includes comments and function calls like 'plt.title()' and 'plt.xlabel()'. Below the code, a scatter plot displays the results, showing a distribution of blue dots representing galaxy positions. The plot is overlaid with a grid and labeled with 'Crab', 'Cyg X-1', and 'Sag X-2' in red text.

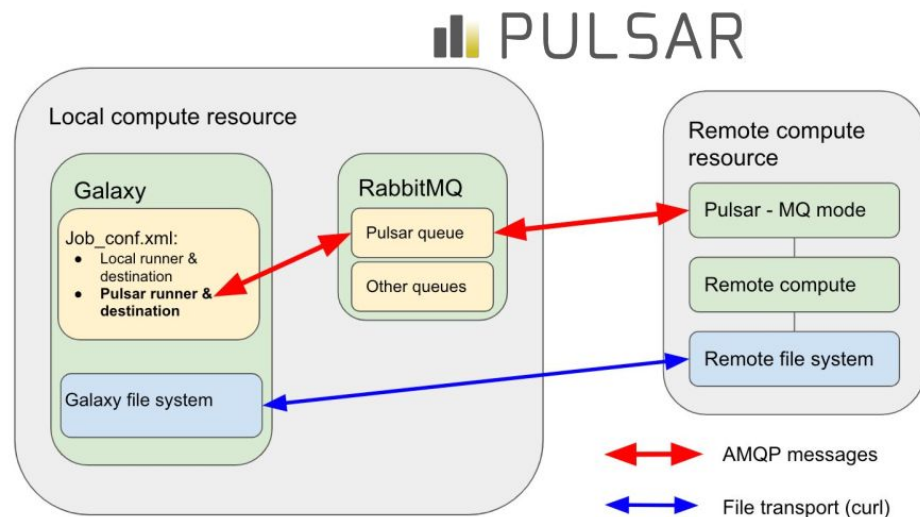
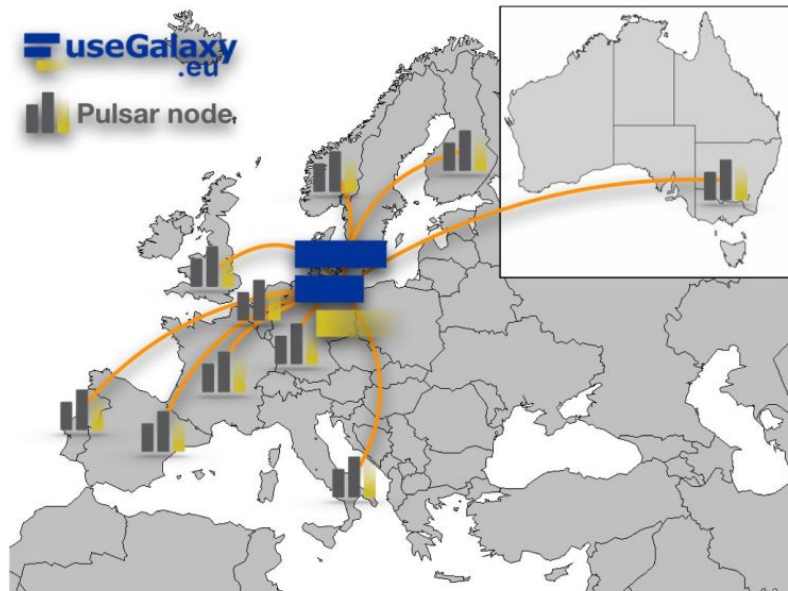
```
# healpy.Mollview(fig=plt.gcf(), number, grid=True)
healpy.graticule()
plt.title("")
# plt.xlabel("RA")
# plt.ylabel("Dec")
# plt.grid()
```

(78): Text(0.5, 1.0, '')

Pulsar distributed compute network

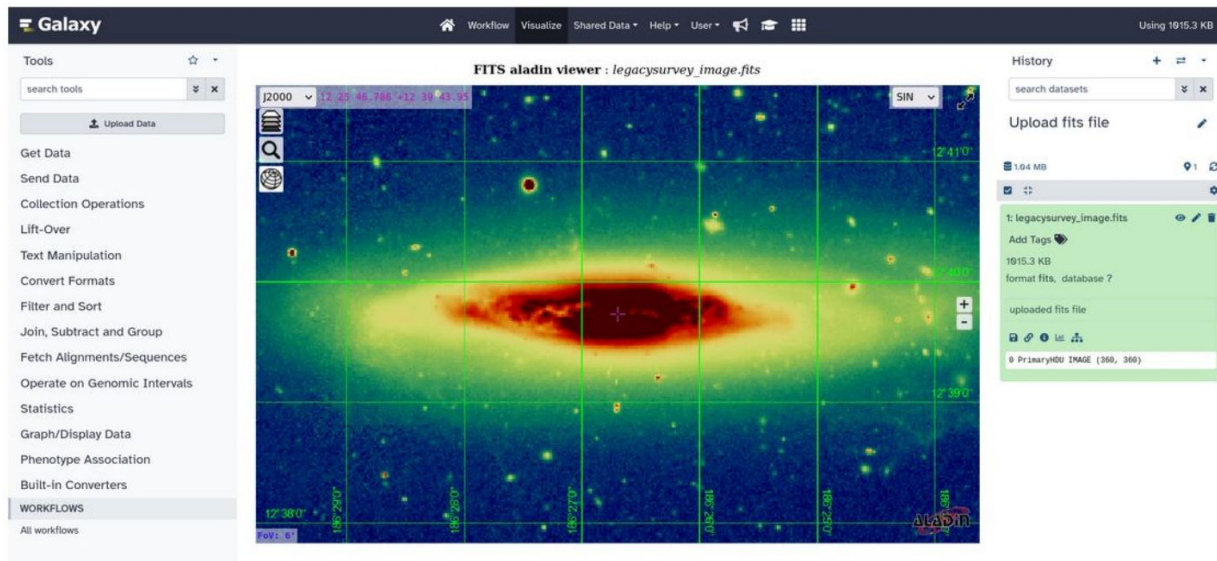
Pulsar services connected in a network enabling large computing network between **European supercomputing centers**

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).



The screenshot displays the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'User', and 'Using 1015.3 KB'. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by a list of tool categories: Get Data, Send Data, Collection Operations, Lift-Over, Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Group, Fetch Alignments/Sequences, Operate on Genomic Intervals, Statistics, Graph/Display Data, Phenotype Association, Built-in Converters, and WORKFLOWS (All workflows). The main area shows the 'FITS aladin viewer : legacysurvey_image.fits'. The viewer displays a color-coded astronomical image of a galaxy with a grid overlay. The grid is labeled with J2000 coordinates: 12 25 46.788 +12 39 43.95 at the top, and 12° 38' 0" at the bottom left. The right side of the viewer shows a 'History' panel with a search bar and an 'Upload fits file' section. The history entry shows '1: legacysurvey_image.fits' with 'Add Tags', '1015.3 KB', and 'format fits, database?'. Below this, it says 'uploaded fits file' and '0 PrimaryHD IMAGE (360, 360)'. The Aladin logo is visible in the bottom right corner of the viewer.

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](#)).

Galaxy Europe

Tools: astro

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

Tool Parameters

Archive Selection
Query specific IVOA archive

Astronomical archives
The VO @ ASTRON TAP service

Query selection
ADQL Query Selection
No specific query (first n files from archive)

Output selection
Number of files or uris to download: 10
Beware of disk space usage when downloading large number of files!

Tool output type selection - optional
 Select / Deselect all
 Return URL list as CSV
 Download files
 Return URL list in extended HTML (requires HTML rendering permission, see help)
 Return URL list as HTML

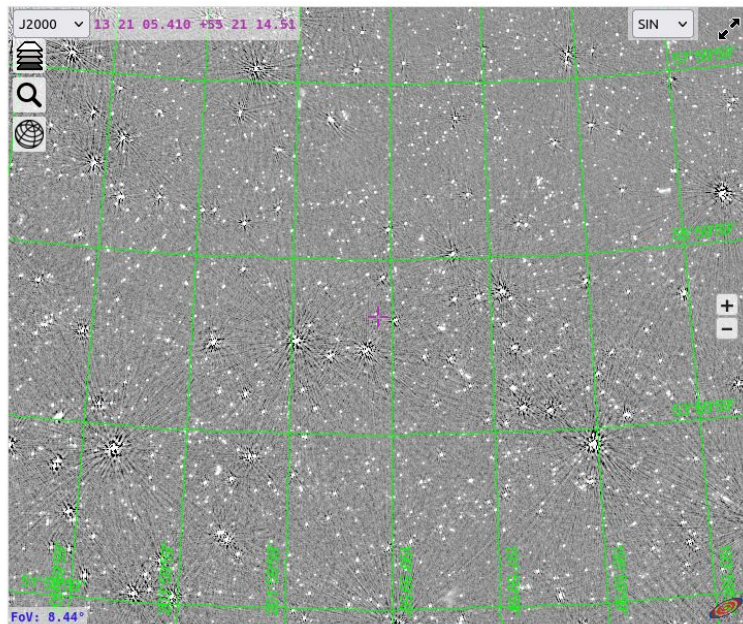
Additional Options
Email notification
 No
Send an email notification when the job completes.

Run Tool

SOURCES PREVIEW ARCHIVE: <https://vo.astron.nl>

ivoa.obscure

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/~
3	apertif-dr1	190807041	190807041_AP_B015_0	ivo://astron.nl/~?APERTIF_DR1/190807041_AP_B015/HI_imag
3	apertif-dr1	190807041	190807041_AP_B015_1	ivo://astron.nl/~?APERTIF_DR1/190807041_AP_B015/HI_imag



15.0 KB
format html, database ?

100% CPU

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 KB
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

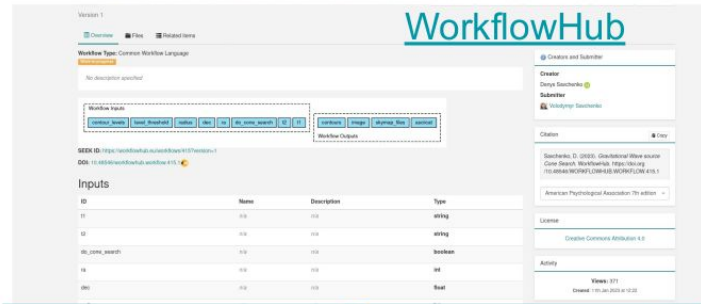
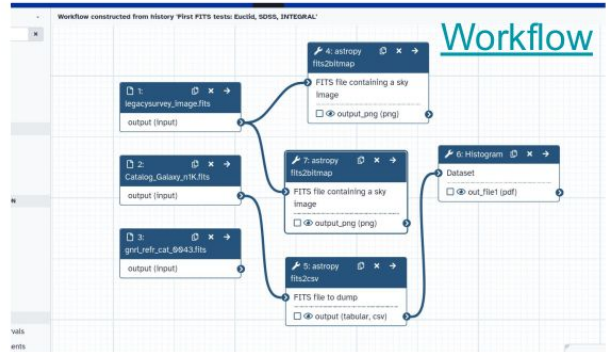
Stored in discoverable workflow catalog:  WorkflowHub

We are connecting tool catalogs by ingesting tools developed in  SDSC RenkuLab and with  MMODA (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. →

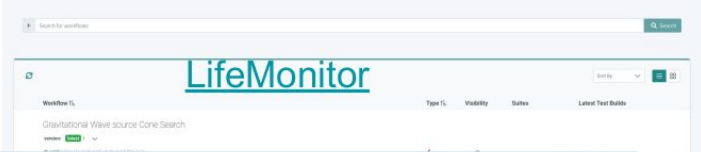


The screenshot shows the WorkflowHub interface. It features a search bar, navigation tabs (Overview, Files, Revisions), and a workflow card. The card includes a description, a table of inputs, and metadata.

ID	Name	Description	Type
11	11	11	string
12	12	12	string
13	13	13	boolean
14	14	14	int
15	15	15	float



The screenshot shows a dashboard with four colored tiles: a teal tile with '18', a yellow tile with '0', a red tile with '7', and a grey tile with '2'. There are also icons for 'Tools pending', 'Failed', and 'Personality'.



The screenshot shows the LifeMonitor interface. It features a search bar and a table of workflow statistics.

Workflow ID	Type ID	Visibility	Status	Latest Test Builds
Gravitational Wave source Cone Search				



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**