MMODA multi-messenger data analysis platform

Denys Savchenko, A. Neronov, V. Savchenko, C. Ferrigno, G. Barni, M. Meharga, O. Kalashev

François Arago Centre

EPFL

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

Abstract

MMODA (Multi-Messenger Online/Open Data/Distributed Analysis) is an astroparticle community initiative carried out in partnership between François Arago Centre (FACE) of the APC laboratory, the University of Geneva and EPFL. **MMODA** aims to provide **a real-time, online data analysis platform.** The platform makes use of **cloud-based** data management solutions and virtualization technologies. It is intended to meet the challenges of efficient **sharing and reuse of data and data analysis workflows** and the **reproducibility of results.**

Apart from standalone analysis plugins, we enable an **automatic deployment** of the workflows in the form of **semantically-annotated** Jupyter notebooks. This process enables efficient crowdsourcing of the **analysis-as-a-service** modules encompassing the domain expertise.

Motivation

Multi-messenger astronomy is an exploding field!
Last decade, key new observables were discovered, and conventional telescopes dramatically upgraded to match. The **number of alerts and volume of data** we deal with **increased by a couple orders of magnitude** in the last years, and several nearly-ready telescopes promise another comparable increase.
A wealth of astronomical sources emits particles over **a very broad energy range**.

Understanding the emission mechanisms requires data collected with many types of telescopes. Bunch of them appears on the sky for **a short period of time,** thus require fast reaction to observe them with different types of telescopes while they are active.
Individual astronomers cannot master data analysis techniques of all these telescopes at once.
A system that extracts **analysis-ready results** automatically would be useful.

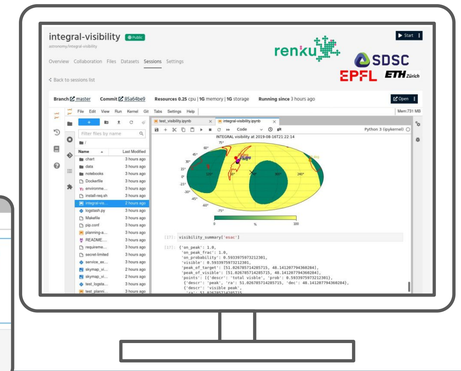
Development space

Creating analysis services is a complicated task.
Need expert astronomers with state-of-the-art tool-building skills. There are much **more scientists who can make a Jupyter notebook than write organized code.** There exist a bunch of Jupyter-based online data-analysis services. **JupyterHub(s), Google-collab, ESA DataLabs, Renku** etc.

We collaborate with Swiss Data Science Center to incorporate the **Renku** platform into the ecosystem.

- **Continuous integration and testing.**
- Support for **publishing of data and code** (e.g. in Zenodo).
- Support **annotating** with semantic terms to simplify reuse by scientists and robots.
- Developing and integrating metadata in **Knowledge Graph.**

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

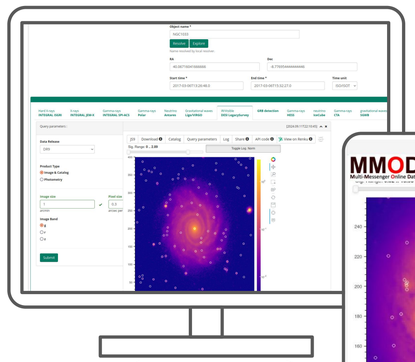


The platform

MMODA Multi-Messenger Online Data Analysis

online, cloud-based tool for exploring, transforming multi-messenger data

https://mmoda.io



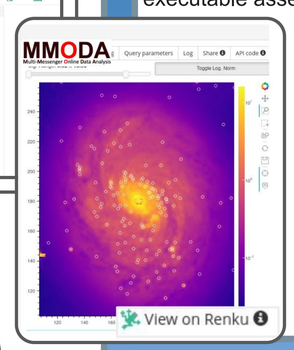
The platform provides access to high-level products like lightcurves, spectra, images by pre-processing (mostly public) observational data of different facilities: X-ray/Gamma space-based instruments, ground observatories ranging from radio- to gamma-, neutrino and GW detectors.
We **leverage workflow as a service**, possibly calling from another workflow. **We are publishing the live tool, not just it's output,** assuring reproducibility and reusability of results.

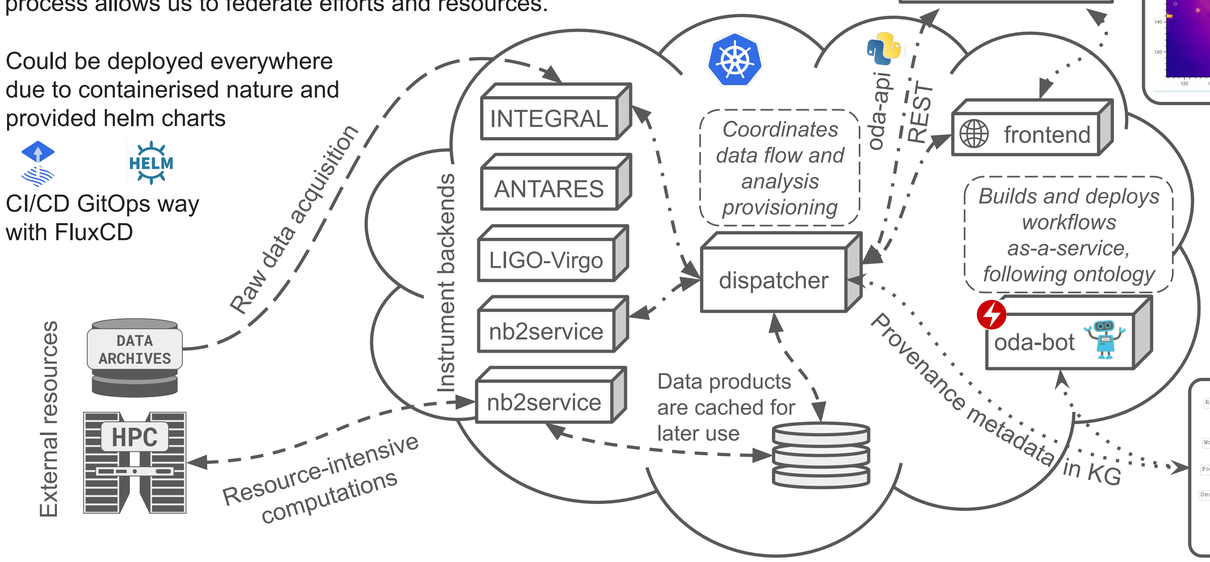
https://github.com/oda-hub

Collaborative open-source community development process allows us to federate efforts and resources.

Could be deployed everywhere due to containerised nature and provided helm charts

CI/CD GitOps way with FluxCD



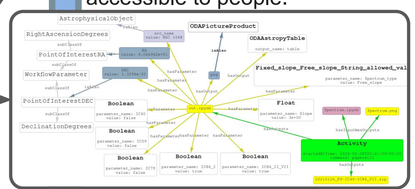


Workflows creating cycle

Linked Open Data Knowledge Graphs: robots gaining creativity

People know a lot, and form free associations. Robots have much information too, but may lack context (general knowledge). E.g. much insight is reported in GCN Circulars but only accessible to people.

- **Global linked identifiers – URIs** (ivo://, http://, …): building common vocabularies. URIs point to documents, workflows, data, astrophysical objects.
- Explain possible **relations** in **ontologies.**
- Embedding and following references, to express connections between different URIs.



KG-informed automation

- **Act on new paper or observation report**
- **Act on new software or data:** re-run analysis of a "test case", ensuring assumptions about instruments
- **Act on new observation:** testing assumptions about physical reality
- **Act on new platform or time moment:** make sure the platform runs smoothly and is sane

Harvest astro-specific metadata from code runtime; integrate code/workflows with **data** by keeping provenance information in the form of **knowledge graph.** Linked Data ideas: make it connected into larger world with its language.