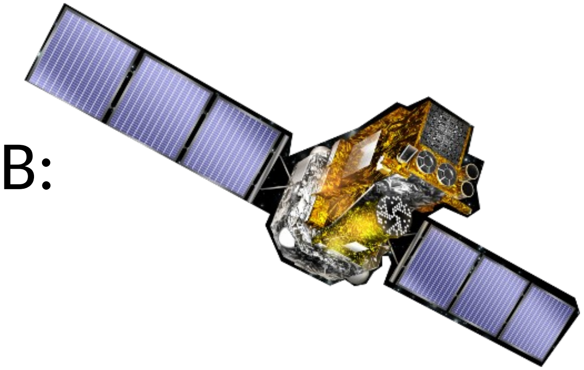
UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'astronomie

EPFL

n|w University of Applied Sciences and Arts Northwestern Switzerland

The AstroOrdas project



swissuniversities

Swiss Open Research Data Grants Track B: Establish projects (PI S. Paltani)

Carlo Ferrigno

Volodymyr Savchenko, Gabriele Barni, Mohamed Maharga, Andrii Neronov, Stéphane Paltani, André Csillaghy, Emma Tolley, Jean-Paul Kneib

University of Geneva

MMODA
Multi-Messenger Online Data Analysis

The landscape

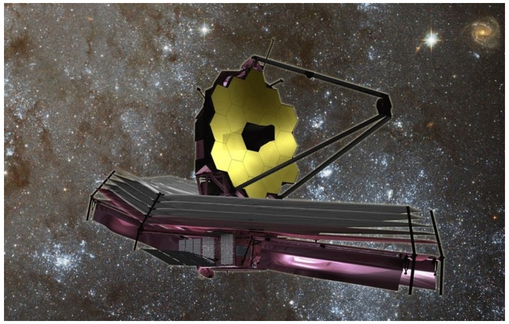
- Astronomical data are often in public archives after a proprietary period
- Their reduction can be cumbersome and require expert users
  - We want to reduce the usage barrier by providing web-based services
- Analyses are complex and workflows are often not reproducible
  - We want to provide a machinery to ease the link between papers and workflows

Use Case

- In 2017, the first joint detection of a gravitational wave and a gamma-ray burst open the path to multi-messenger astronomy
- Many telescopes and facilities had to be exploited fast an reliably: 3000 scientists signed a summary paper.







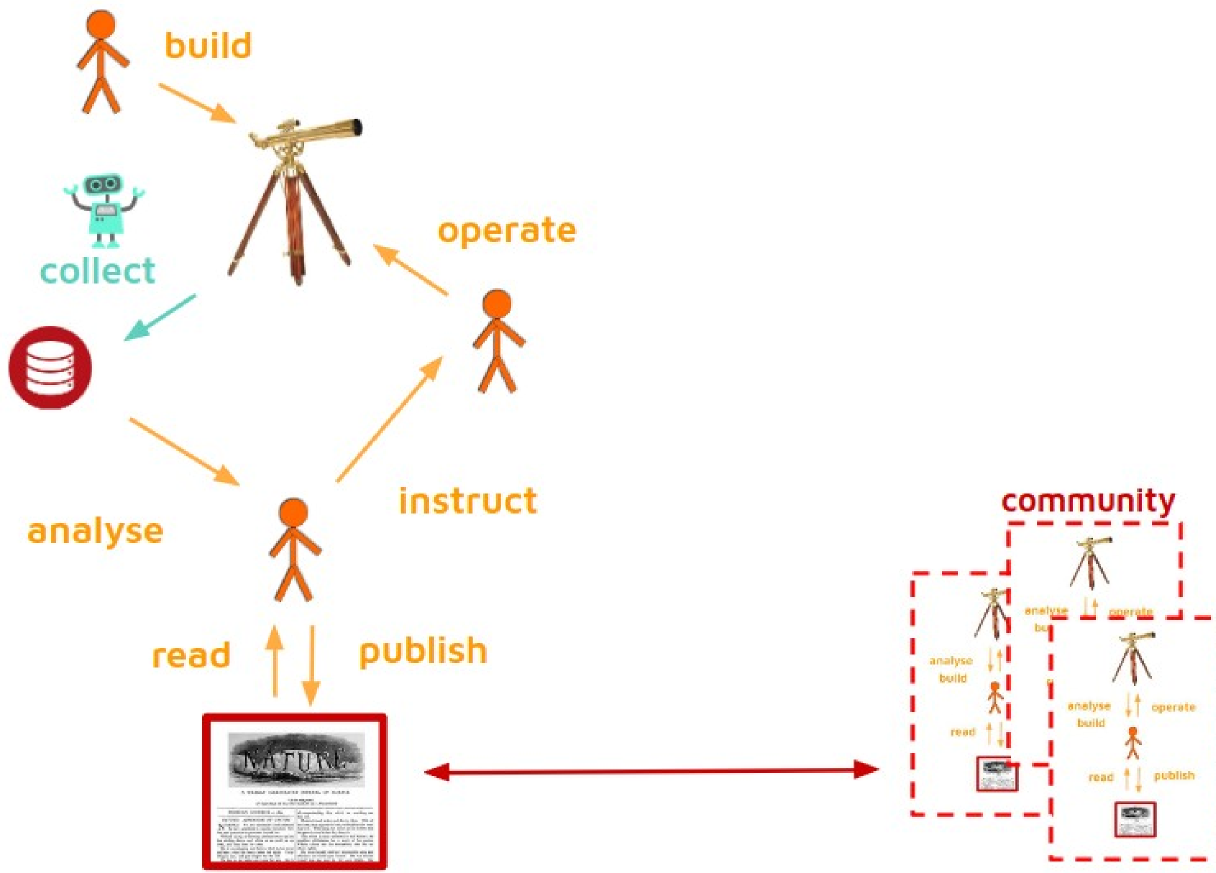
Gravitational wave
Optical
Radio
Cherenkov telescopes
Space telescopes
Neutrino telescopes

MMODA
Multi-Messenger Online Data Analysis

Where we come from

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



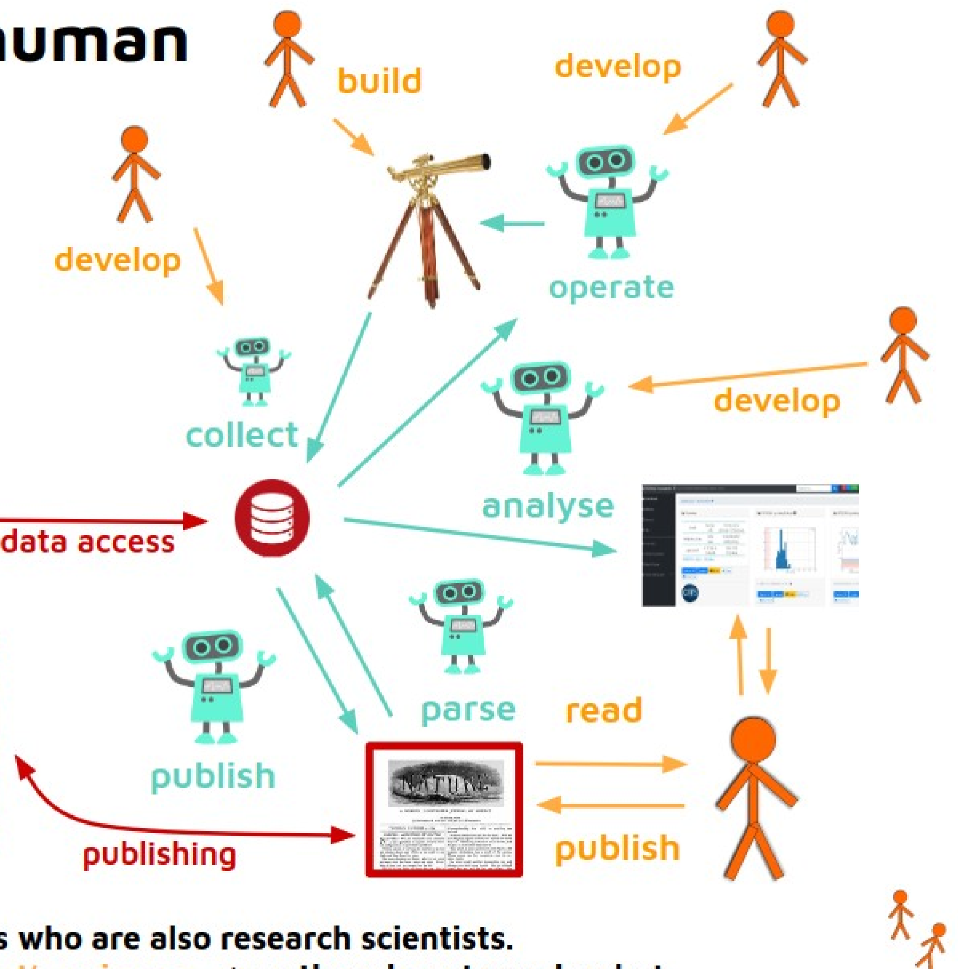
Human reaction and processing is slow, even if it's within even one person. But people are smart

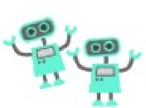
How we evolve

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
Multi-Messenger Online Data Analysis

Specific aims of the project

- Develop an ecosystem of cloud-based services and technologies to provide added value to data from science data centers for astronomy, astroparticle and cosmology projects
- Deliver an accessible implementation of the Findable Accessible Interoperable Reusable (FAIR) principles for Open Research Data
- Connect to the world-wide network of online astronomical ORD analysis services (ORDAS)
- Connect data and algorithms to scientific publications in a viable way

Missions/experiments involved

- We are involved in data centers for:
  - INTEGRAL (X-ray, gamma-rays)
  - Gaia (cornerstone, astrometry)
  - Euclid (Cosmology)
  - CTA (TeV range)
  - SKA (radio)
- Experts can either make full ecosystem for data reduction or develop specific analysis workflows

Collaboration with similar projects

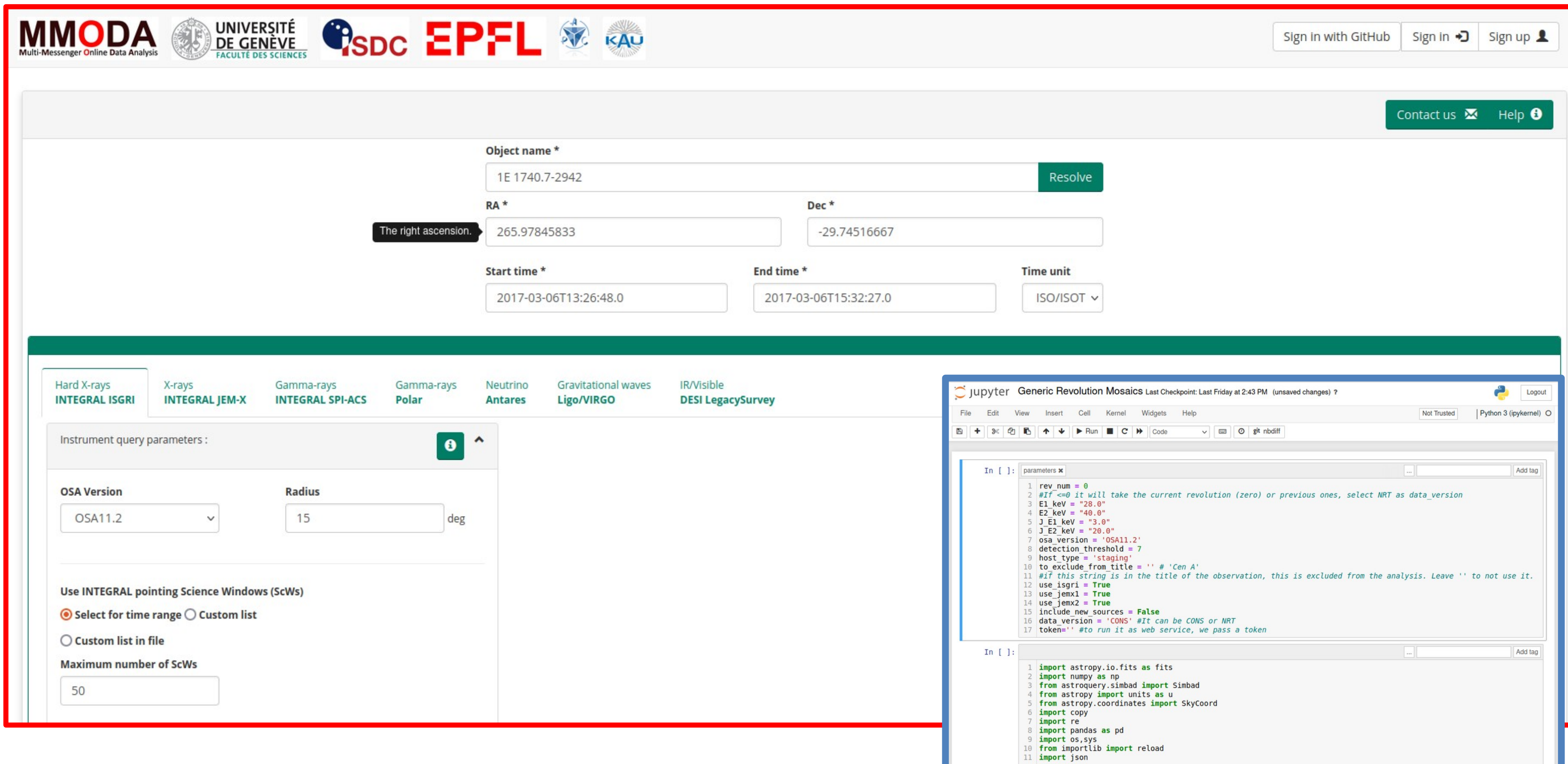
- We exploit resources and make joint development with the SDSC through its Renku platform (see talk)
- We work in close collaboration with Euro Science Gateway with whom we share resources (see talk)
- We also share resources and expertise with the INTEGRAL Science Data Centre at the University of Geneva

MMODA
Multi-Messenger Online Data Analysis

Multi-Messenger Online Data Analysis

Data analysis of telescopes and astronomical messenger detectors can now be done on remotely via dedicated services accessible via

- Web interface (in a browser)
- Application Programming interface (API) from e.g. Jupyter notebooks



MMODA
Multi-Messenger Online Data Analysis

Publicly available workflows

- Images spectra and light curves from two instruments on board of INTEGRAL, used for easy access
- The only source of Gamma-ray bursts reduced light curves from Polar (a Chinese-Swiss mission)
- Desi Legacy survey: sky images and photometry data in the infrared-to-visible band, the photometry data is obtained from the queries of TAP service provided by Astro Data Lab
- Interface to ANTARES public data from with the directions of arrival of events detected by the ANTARES neutrino telescope and used for point source searches from 2007 to 2017

MMODA
Multi-Messenger Online Data Analysis

A flexible platform

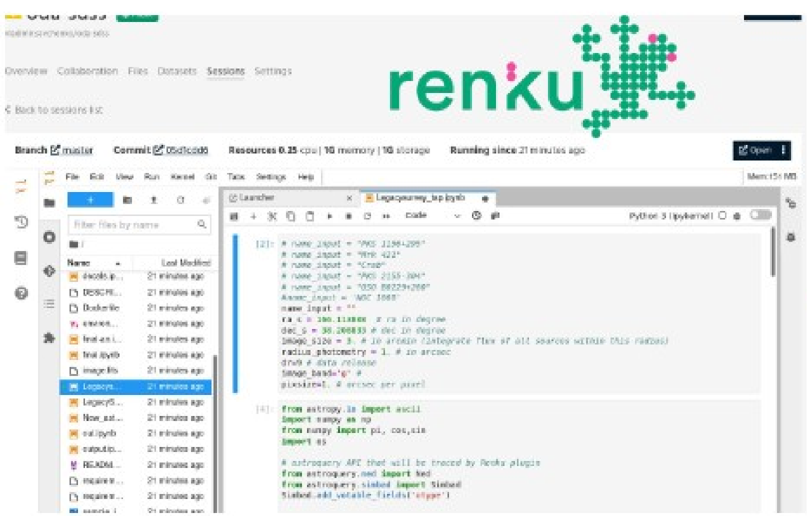
- We managed to federate resources from France to Ukraine to setup the MMODA services by organizing a collaborative open-source community development process in GitHub https://github.com/oda-hub
- This provides scientific added value for the astronomical ORD.
- These online services can now be extended in a modular fashion thanks to the RENKU platform (SDSC) to include, as a first priority, the data sets of the astronomical telescopes and space missions in which we are involved.
- Services can be deployed anywhere owing to containers

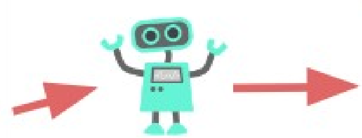
MMODA
Multi-Messenger Online Data Analysis

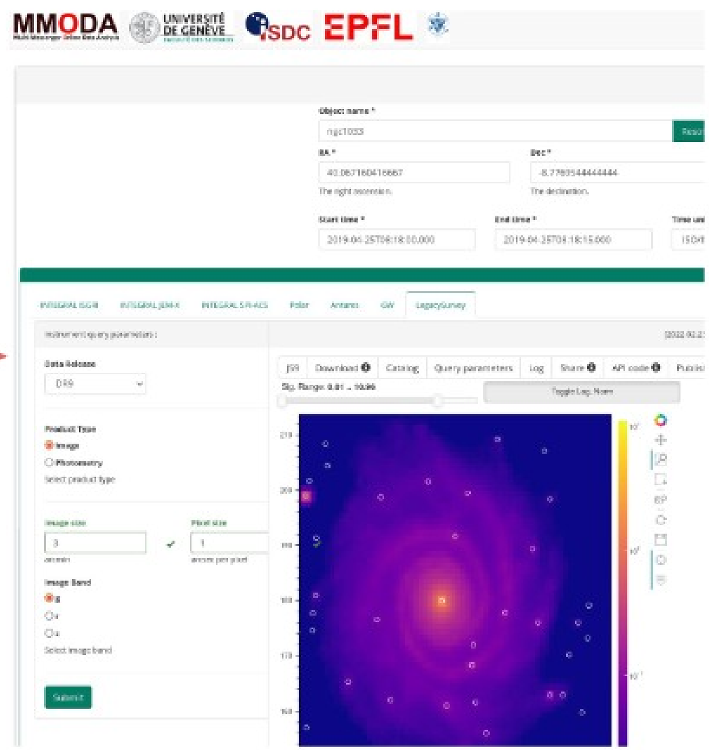
From workflow to web services

- We can use notebooks with input parameters to produce web services.
- This leverages on annotations stored in an extensible knowledge graph that drives the production of a web interface
- Versatile but challenging to customize

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







The MMODA product gallery

- An explorable collection of high-level products from INTEGRAL instruments (for now)
- Hosted on a Drupal-powered website with full REST-API access for content creation, editing, and exploration.
- A legacy archive linked to a flexible tool

MMODA
Multi-Messenger Online Data Analysis

Enhanced Provenance

Notebook

[isgri-full-image-sextractor.ipynb]
Notebook parameters

- The python notebooks used to create these products have been developed by us and are linked at the page of the product together with the input parameters.
- The analysis is fully reproducible -> FAIR and open data !

isgri notebooks
Carlo Ferrigno authored 8 hours ago

isgri-full-image-sextractor.ipynb 22.49 KiB

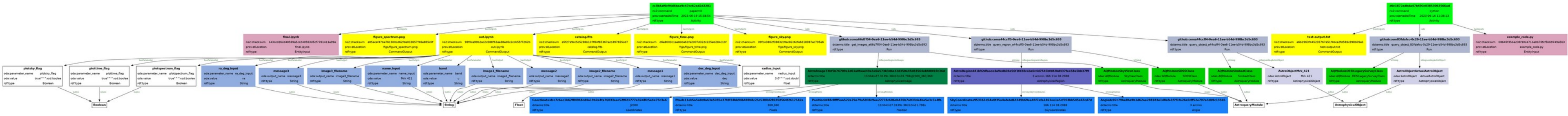
```
#General search
tstart='2003-02-01T00:00:00'
tstop='2022-10-31T23:59:59'
#I use 3.5 deg for JEM-X and 12 for ISGRI, by running twice the notebook.
source_name="3C 273"
radius=10

osa_version="OSA11.2"
data_version='cons'
integral_data_rights='all-private'

#### NB
# This is the limit for each call (to be raised to 500)
s_max=150

#For ISGRI image and light curve
E1_keV=28.
E2_keV=50.
#for image and catalog extraction
detection_threshold=7.0
```

- Needed an account in MMODA to reproduce the analysis
- Upload permission to the gallery is restricted, but it can be granted.

MMODA
Multi-Messenger Online Data Analysis

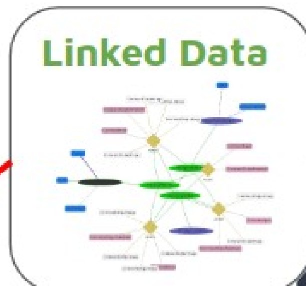
Graph visualization

- The ontology describes relations between some things, terms. We develop our own ontology leveraging on existing ones as Virtual observatory to annotate workflow input and outputs.
- This allows us to build transformation graphs in which terms are linked by relations.
- We catch calls to external astronomical services and made a tool to visualize them.
- Full provenance is displayed



Connect papers and workflows

- If workflows can be shared and served, we can cite them in papers.
- We work in connection to a Knowledge graph of workflows (challenge of size and computing power)
- We can embed results directly in papers.




```
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}
```

https://github.com/oda-hub/linked-data-latex/

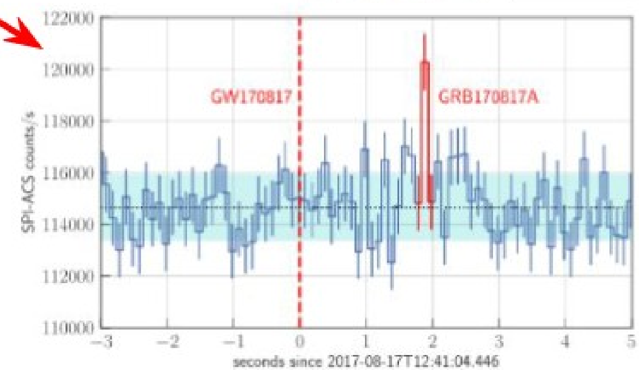


FIG. 2.— SPI-ACS light curve of GRB 170817A (100 ms time resolution), detected 2 seconds after GW170817. The red line highlights the 100 ms pulse, which has an S/N of 4.6 in SPI-ACS. The blue shaded region corresponds to a range of one standard deviation of the background.

MMODA
Multi-Messenger Online Data Analysis

SKA - SRC

We plan to connect the SRCNet to AstroORDAS

- Based on the data cubes or visibility from the data challenge, we will create a service to query the data set from the interface.
- We will create metadata for the SDC3 data challenge and store them in a separate DB.
- General task: use MWA data as precursor for testing access to SRC data
- MWA data will be used in SKACH science use cases to test functionality necessary for accessing and processing data. We will make sure that the use case can also be accessed and integrated in the MMODA environment.
- Possibly make the connection to the MWA data processing.
- Include the results in the user interface of MMODA.

MMODA
Multi-Messenger Online Data Analysis

Technical development lines

- Develop specialized visualization tools for astronomical products. Open source is a key asset to make it useful.
- Develop extension to knowledge graph visualization and contribution. This helps consuming and sharing the knowledge on the network and make flexible visualization.
- Diffuse technology to link products to papers, which is now cumbersome
- Create a way to deploy easily heavy computation on HPC clusters so to easily federate resources.

MMODA
Multi-Messenger Online Data Analysis

Conclusions

- Swiss Universities has recognized the value of not only providing open data, but also open analysis services
- We aim to provide FAIR data using **open-source technology** that could be exploited by other partners and facilitates data reduction and reproducibility of results.
- We are federating in Switzerland and outside Switzerland. This enhances the possibility of sustaining the project on the long term, despite the difficulties in finding reliable funding resources.
- We have connections with many projects, but we are open to collaboration.