UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'astronomie

Workflow interoperability

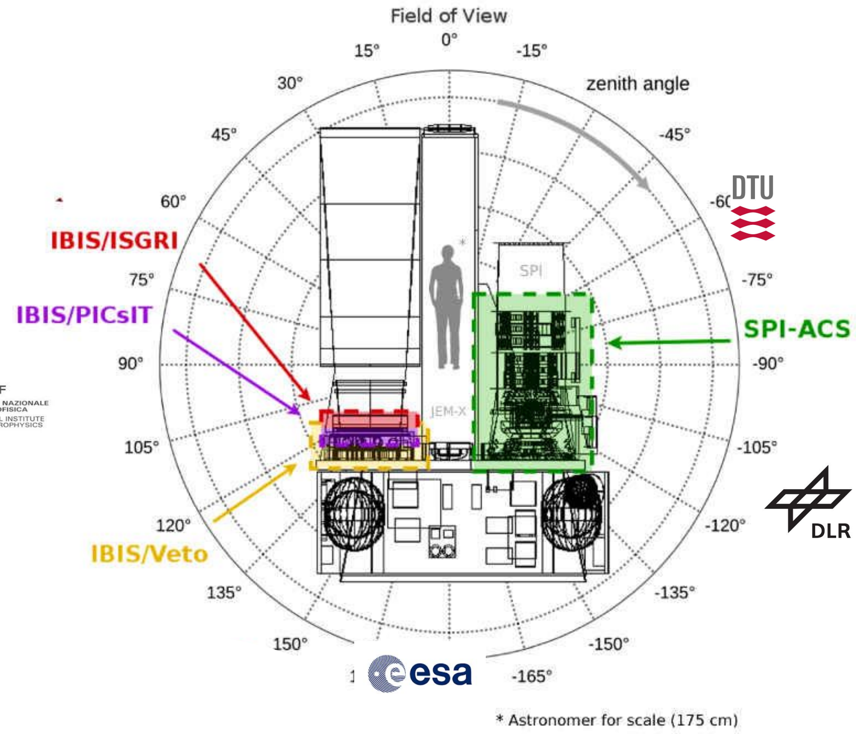
for telescope operations and time domain astronomy

Volodymyr Savchenko

UNIGE

IVOA May 2021 Interoperability Meeting
27 - 05 - 2021

INTEGRAL space observatory Data Center, 2002+



Core role of **ISDC** is to provide community with means to leverage the telescope, we:

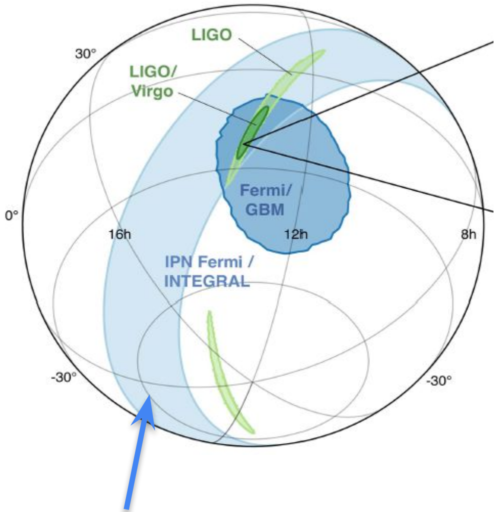
- Consolidate and provide **data, software, services**.
- Perform **quick-look analysis**, especially relevant in time domain astronomy
- Keep an eye on **instruments**

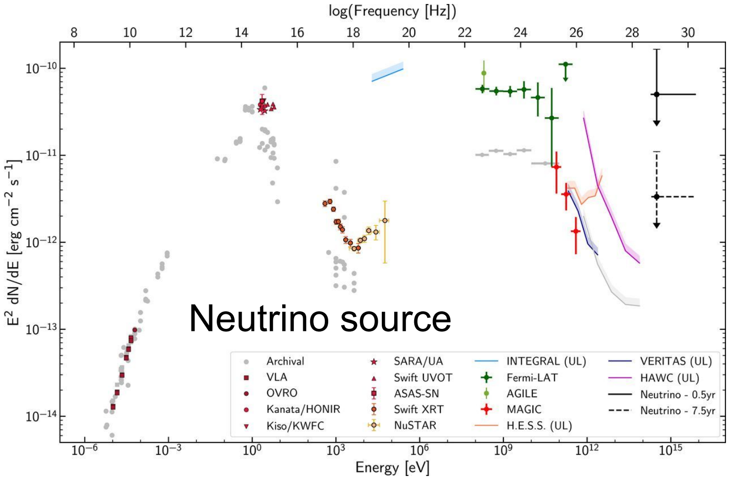
INTEGRAL provides best results by **combining with spectral, timing information from other missions**, and to make an impact on their own, it was decided to **provide API to access** public data, instead of hiding data/workflows and sending uninformative **publications**.

This mean that we need to **handle FAIR live services/APIs/workflows** not just **FAIR data**.

VS+ 2017

Multi-messenger time-domain astronomy is collaborative





Fermi + INTEGRAL Triangulation

unique multi-mission approach

Our focus on **broad synergies** allowed us to take a leading role or contribute in some of the key recent discoveries in our domain:

- Detection of the **first Gravitational Wave - Light** coincidence (2017)
- First detection of light emission from **high-energy neutrino source** (2018)
- Discovery of the source of mysterious **Fast Radio Burst** (2020)

VS+ 2017, LVC 2017

Traditional INTEGRAL transient analysis

Research, development environment lets experts develop, test, and integrate:

- data reduction
- theoretical models
- Spacecraft operation tools

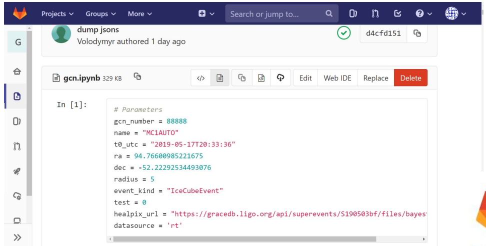
Observers and Operators:

- **Find combinations** of data, adapters, statistical methods, publishers, planners
- **suggest new observations**
- **distribute** results

experts

data
software
documents

Observers,
Operators,
Shift



GitLab

Nature events
VOEvent, HTTP, Kafka, etc

```
TITLE:   GCN CIRCULAR
NUMBER:  25505
SUBJECT: LIGO/Virgo S190828l:
in INTEGRAL SPI-ACS prompt ob
DATE:    19/08/28 08:59:07 GM
```

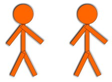
Not good enough if we want to try many combinations quickly: be exhaustive and ready for unexpected
Setup automation for low latency and high-variety activity.

“Standard” modern INTEGRAL transient analysis

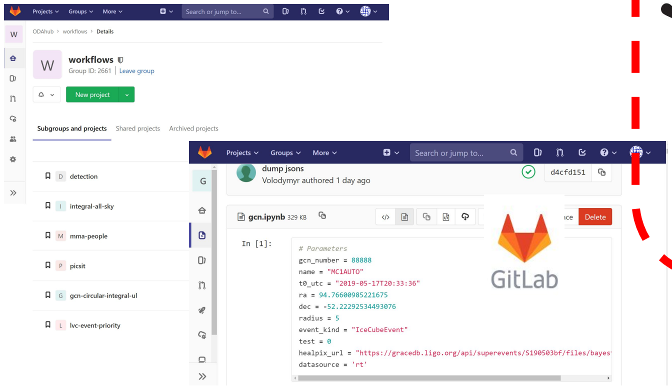
Research, development environment lets experts develop, test, and integrate:

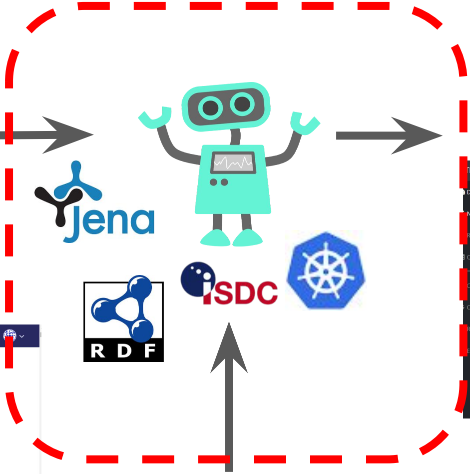
- data reduction (close to data)
- theoretical models (linked to literature)
- statistical methods (as portable as possible)
- Spacecraft operation tools

- **Find combinations** of data, adapters, statistical methods, publishers, planners
- **suggest follow-up**
- **distribute** standard results with public data, uploads to zenodo sandbox.



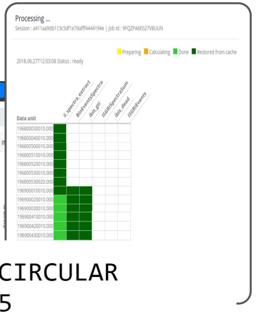
experts

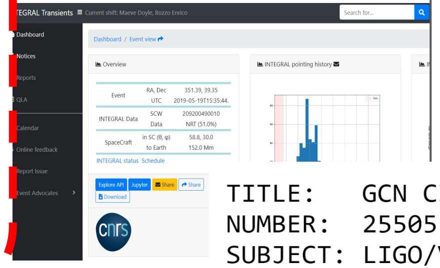




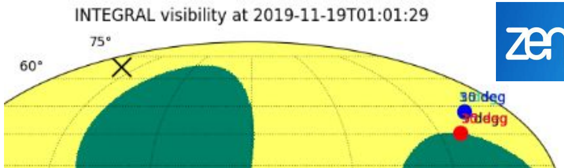
VOEvent, GCN, ATel, Kafka, etc

Shift (24/7)





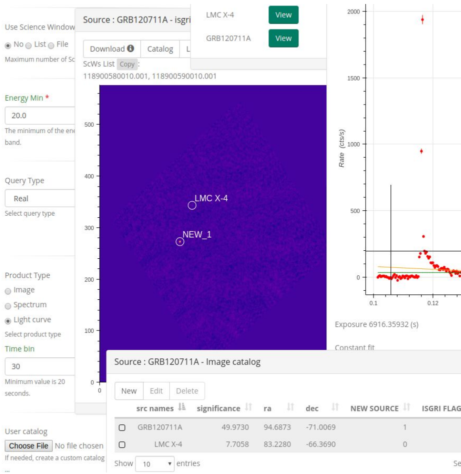
TITLE: GCN CIRCULAR
NUMBER: 25505
SUBJECT: LIGO/Virgo S190828l: in INTEGRAL SPI-ACS prompt ob
DATE: 19/08/28 08:59:07 GM

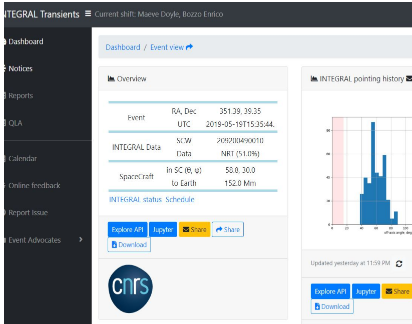


Addressing our challenges relies on **FAIR** Work Space

To optimize collaborative efforts, in this environment, we need robust means of **Finding, Accessing, Interoperating, and Reusing** our assets:

- Documents
- Data
- Software:
    - Code: **scientific software which we build and distribute**
        - Official instrument software (source, binary, doc, more recently containers)
        - INTEGRAL **Quick Look Analysis**
    - Web-based Software (services): **API's and Frontends, which we provide to the community**
        - **Data browse** interface (developed by NASA, hosted un Geneva/Versoix, since ~2002)
        - Help desk, **issue** handling and resolution
        - **Realtime data** interoperability (since 2011)
        - **AstroODA** online analysis (internally since before 2017, first public in 2019)
            - And all of it's **backends** separately
        - **Multi-Messenger Transient Analysis** (since 2018)
        - Various smaller API's for specific purposes





New tools: more and more web-based data analysis

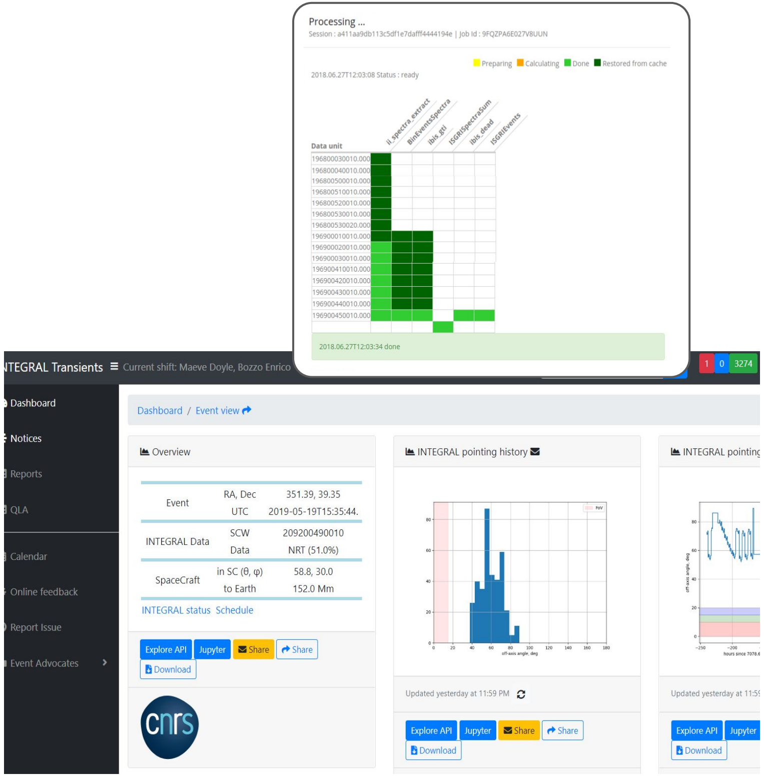
Last years, we develop and consume more and more web-based data analysis services.

In development, we:

- rely on cloud-native technologies, gitlab/github, CI/CD
- Allow domain experts to contribute easily, with **code-to-service** workflow
- support of right-to-replicate

Of course, we use VO services, but also many others. In reuse, we:

- Strive to provide provenance, as much as possible in standard formats (like PROV-O)



https://github.com/oda-hub/ https://www.astro.unige.ch/cdci/astrooda_/

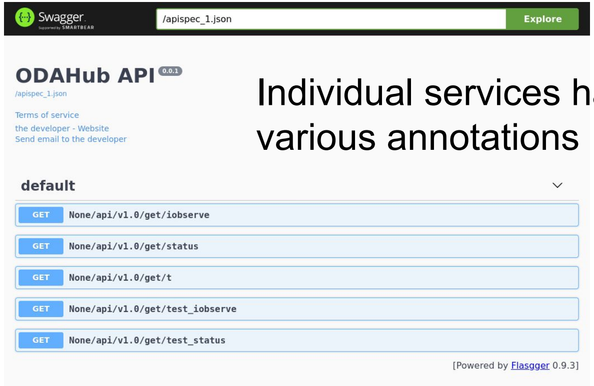
Neronov, Savchenko+ 2021, A&A in press

Sharing workflows is hard

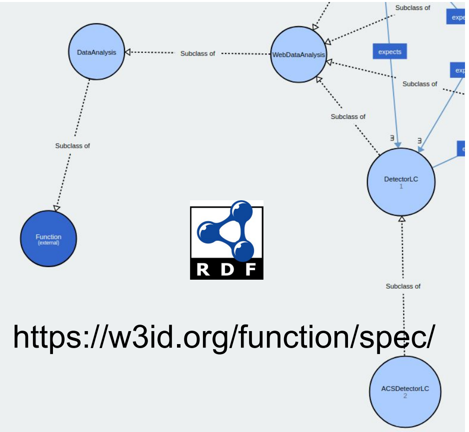
It is possible to annotate and share software like data (e.g. on github or ASCL).

But this would leave too much work to actually **make the software do it’s thing: transform data**.

Without going too ambitious, it should be possible to find a good way to describe collection of workflows (software code and software as a service), describing **software as functions**



Individual services have various annotations



https://w3id.org/function/spec/

INTEGRAL/ODA Knowledge Base (and Graph - **KG**)

Ontology of **processes** based on **fno** when feasible: most important to define input and output formats (types), this allows composition.

Bring some domain context by using **IVOA RDF** vocabularies for **scientific terms and concepts** and **relations** between, as much as needed by our case: a lot of specific things, some general with common terms (e.g. Crab_pulsar is a neutron star) when possible.

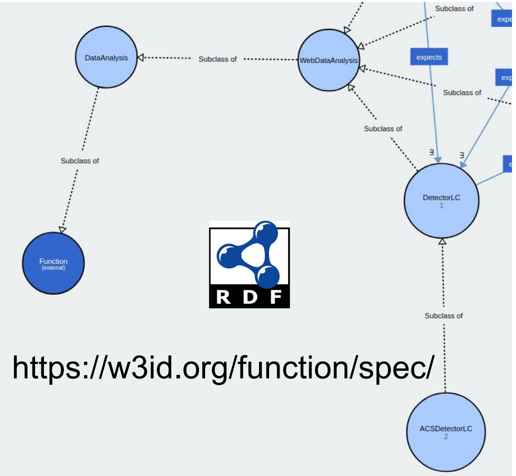
Use Dublin Core, schema.org, whenever feasible.

OWL2 and **SHACL** when possible. But also not always over-obsess over definitions, turns out sometimes ontology can be defined and enforced as KG evolves.

Simple domain-specific **literature parsing** ingests new events in the KG.

Documents, **analysis results**, especially **workflow** executions, ingested and annotated.

Human actors are explicit focus of the KG.



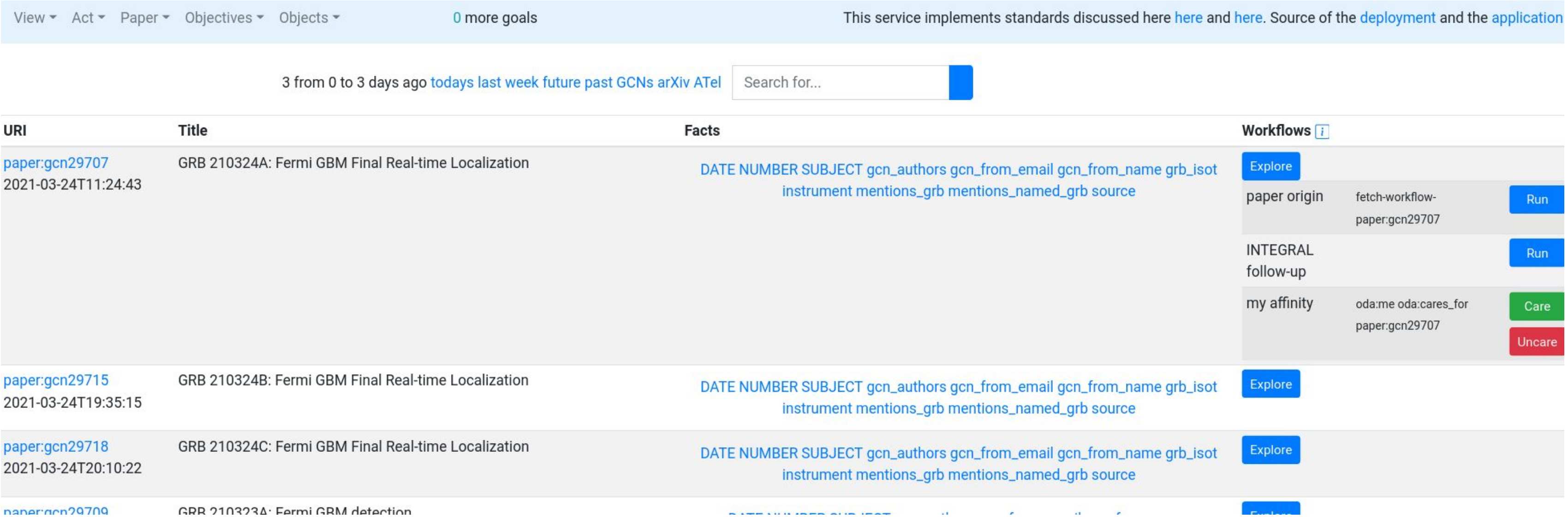
ODA Computational “Experiments”

Since KG contains records of workflows, with I/O types, and data, it is easy to run computational “**experiments**”: combine workflows with data and see what it gives.

Processes that do compositions, *and* ***objective measures*** *are also registered workflows.*

Real classes of compositions:

- “**Act on new paper or observation report**”
- “**Act on new software or data**”: re-do analysis of a “test case”, ensuring assumptions about instruments - **cross-calibration use case**
- **“Act on new observation”:** testing assumptions about physical reality try to find unexpected
- **“Act on new platform or time moment”:** make sure platform runs smoothly and is sane



Growing federated service discovery landscape

EOSC allows to publish and annotate service offerings, helping **live service interoperability.**

ESA currently actively develops similar approach to scientific asset stewardship: **ESA DataLabs** (and we are involved in it as early adopters), with a **science application (~workflow) discovery hub.**

Swiss Data Science Center (SDSC) addresses some of the keys needs valuable for further developments, and has a **Knowledge Graph**, so we could interoperate with it.

We have an active project with **SDSC** to automatically derive workflows annotations in **RDF** from tracking **astroquery** interactions.

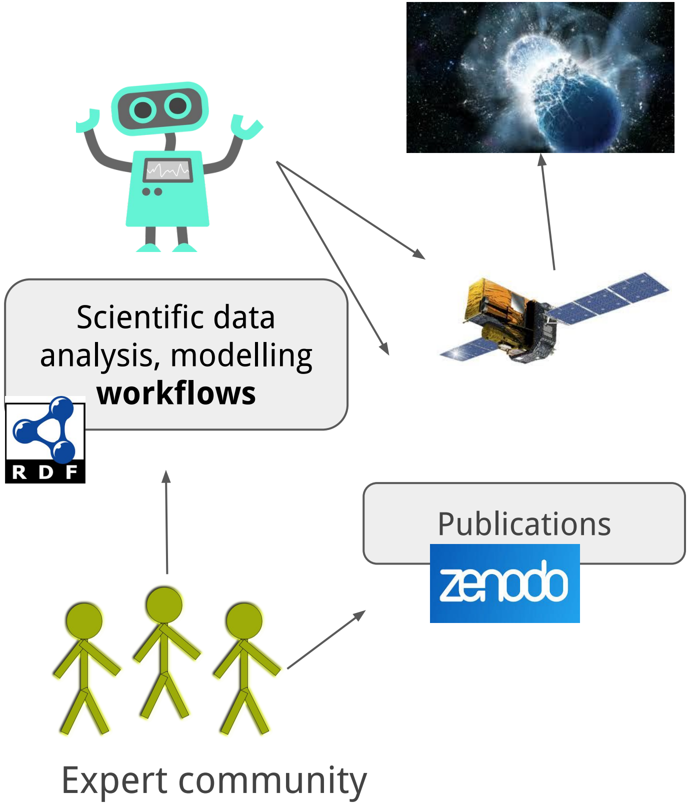
Summary

- Time domain astronomy and telescope operations really rely on “**keeping in touch**” with instruments, space, technologies.
- Growing number of tools, many of them federated, benefits from **more structured annotation**, enabling automated coordination and **composition of data and operations on data.**
- Seems like RDF (despite a number of difficulties) is not a bad way to describe and share this rather diverse assert collection.

Problems, Questions

- Dealing with diversity of languages
  - Keep consistent own world(s) of facts
- Developing ontology is hard. More tools are needed
  - we relied on “just ingesting” and then frequent refactoring, by specifying reasoning rules. It becomes part of the natural graph evolution. But more tools are needed to support this workflow.
- Defining what is authoritative claim
  - Who has a right to make propositions/claims? There are mechanisms with named graphs. We also tried RDF*.
  - Atomic Data restriction on RDF can be useful
- How to adequately share it
- All these mitigations make the **KG very verbose.**
  - Trying to use different graph views and subset to deal with it

Workflow standards let researchers build the system



- Workflow standards foster understandable, usable **publishing** of **Findable Accessible Interoperable Reusable** methods
- Workflows can offert **adapters for data formats**
- Deriving **data, result provenance**
  - explain data
  - trace history
  - credit and blame creators

Provenance RDF PROV-O