

Workflow interoperability

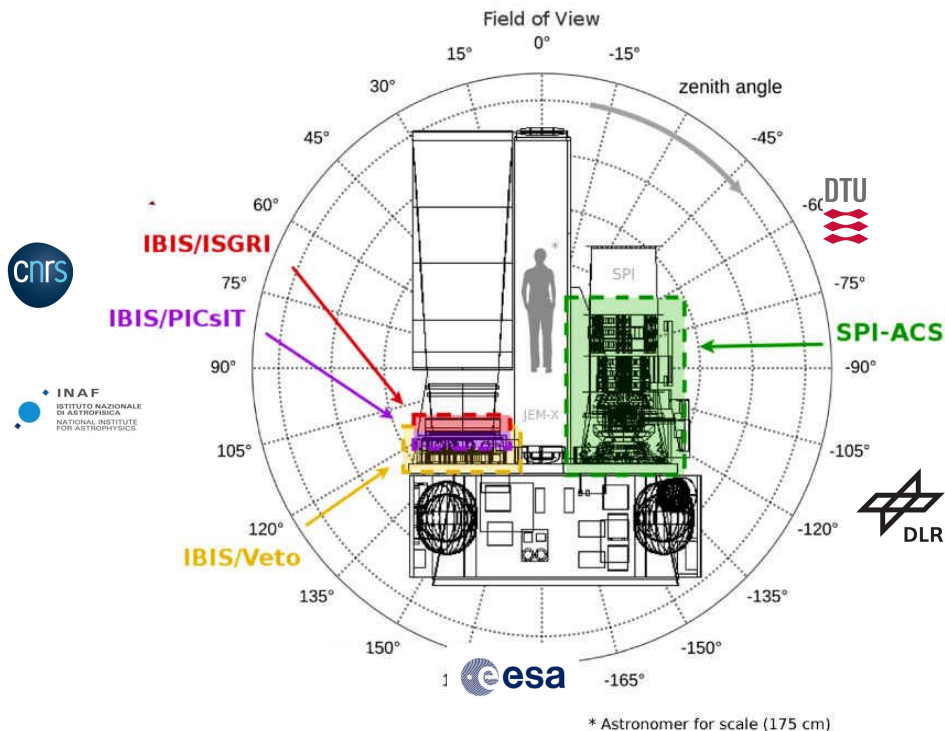
for telescope operations and time domain astronomy

Volodymyr Savchenko

UNIGE

IVOA May 2021 Interoperability Meeting
27 - 05 - 2021

INTEGRAL space observatory Data Center, 2002+



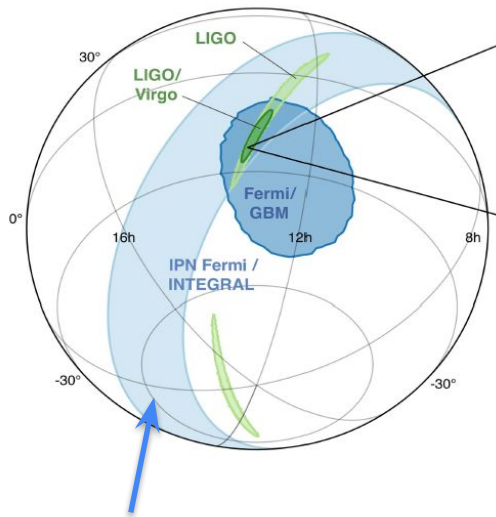
Core role of **ISDC** is to provide community with means to leverage the telescope, we:

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

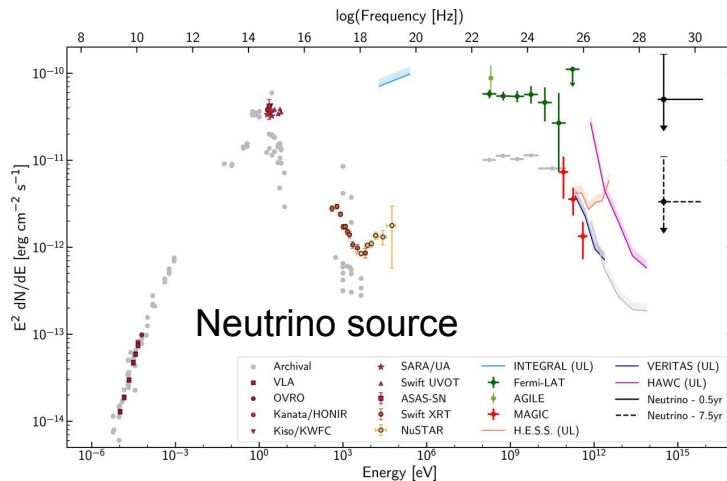
INTEGRAL provides best results by **combining with spectral, timing information from other missions**, and to make an impact on their own, it was decided to **provide API to access public data**, instead of hiding data/workflows and sending uninformative **publications**.

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

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)

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

```
dump jsons
Volodymyr authored 1 day ago
d4cf4151

gcn.ipynb 129 KB
Edit Web IDE Replace Delete

In [1]:
# Parameters
gcn_number = 88888
name = "IC18070"
tb_rtc = "2018-05-17T20:13:13Z"
ra = 54.766089821675
dec = -52.2229254493076
radius = 5
event_kind = "IceCubeEvent"
test = 0
healpix_url = "https://gracedb.ligo.org/api/superevents/S1908281/files/bayes"
dataource = "rt"
```



Nature events
VOEvent, HTTP, Kafka,
etc



Observers,
Operators,
Shift

TITLE: GCN CIRCULAR
NUMBER: 25505
SUBJECT: LIGO/Virgo S1908281:
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:

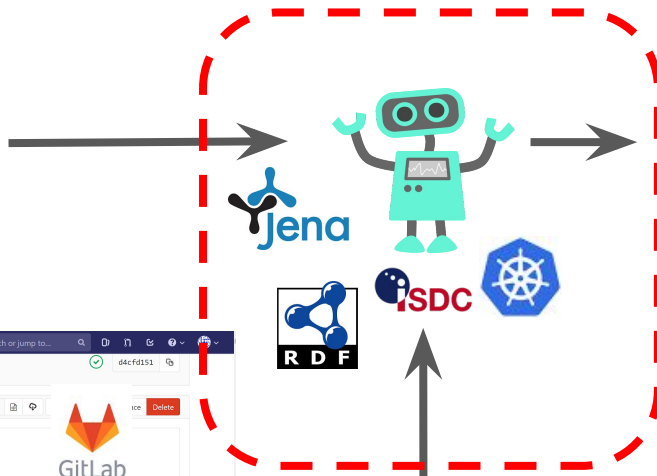
- data reduction (close to data)
- theoretical models (linked to literature)
- statistical methods (as portable as possible)
- Spacecraft operation tools

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



experts

```
In [1]:  
# Parameters  
gcn_number = 88888  
name = "MCIAUTO"  
to_utc = "2019-09-27T20:33:36"  
ra = 342.7608098221075  
dec = -52.22292534493876  
radius = 5  
event_kind = "IceCubeEvent"  
task = 8  
healpix_url = "https://gracedb.ligo.org/api/superevents/S1908281/files/bayes"  
datasource = "rt"
```



Event	RA	Dec	RA J2000	Dec J2000
INTEGRAL DATA	342.76	-52.22	2019-09-27T20:33:36	

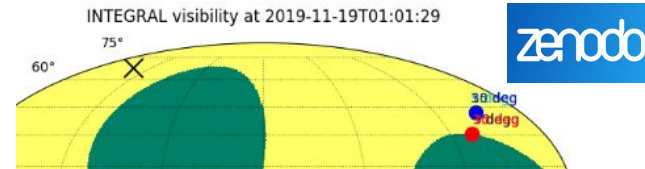
INTEGRAL Global Schedule

Processing - [Event view]

Event	RA	Dec	RA J2000	Dec J2000
INTEGRAL DATA	342.76	-52.22	2019-09-27T20:33:36	

TITLE: GCN CIRCULAR
NUMBER: 25505
SUBJECT: LIGO/Virgo S1908281:
in INTEGRAL SPI-ACS prompt ob
DATE: 19/08/28 08:59:07 GM

VOEvent, GCN, ATel,
Kafka, etc

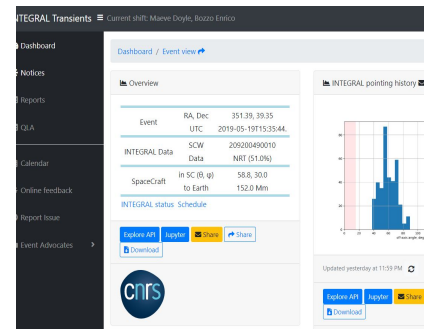
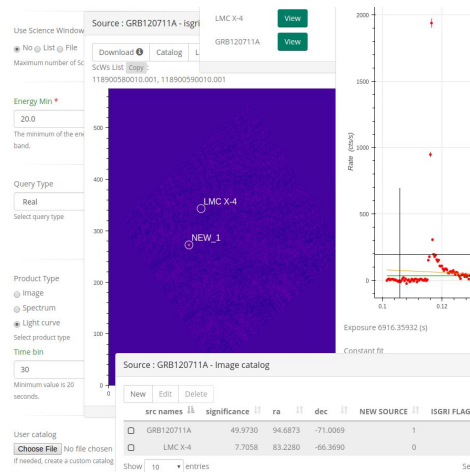


Addressing our challenges relies on **FAIR** Work Space



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, 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

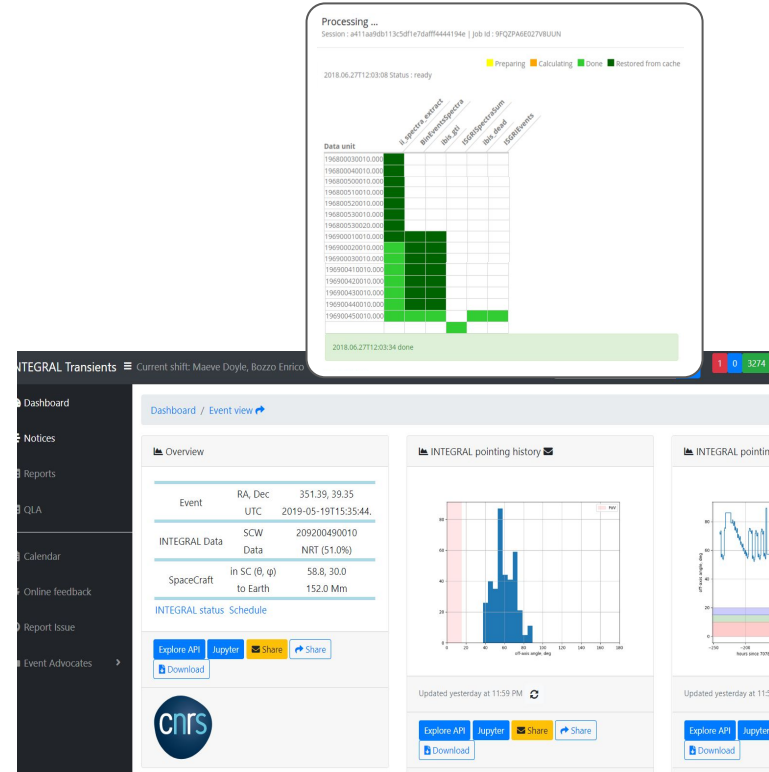
Last years, we develop and consume more and more web-based data analysis services.

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/>

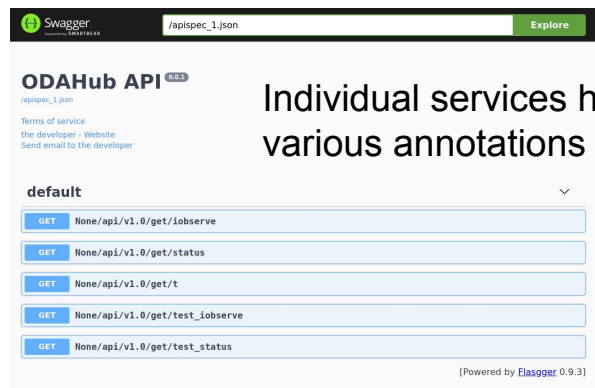
Neronov, Savchenko+ 2021, A&A in press

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**

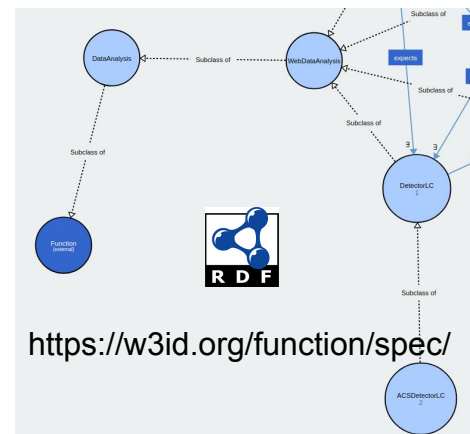


The screenshot shows the Swagger UI for the ODAHUB API (version 0.0.3). The search bar contains "/apispec_1.json" and the "Explore" button is highlighted. The main content area displays the API title and a list of endpoints under the "default" group. The endpoints are:

- GET None/api/v1.0/get/observe
- GET None/api/v1.0/get/status
- GET None/api/v1.0/get/t
- GET None/api/v1.0/get/test_observe
- GET None/api/v1.0/get/test_status

At the bottom, it says "[Powered by Swagger 0.9.3]".

Individual services have various annotations



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.

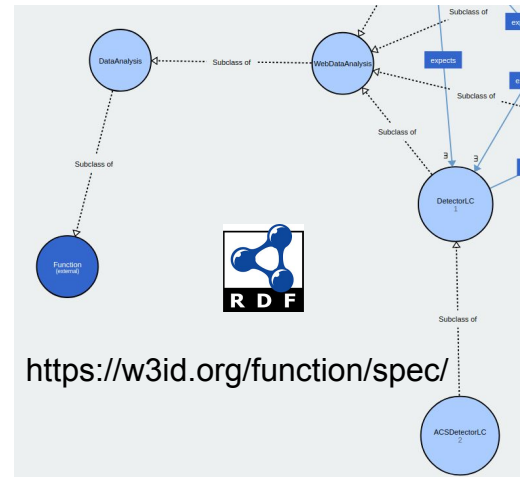
Use Dublin Core, schema.org, whenever feasible.

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

View ▾ Act ▾ Paper ▾ Objectives ▾ Objects ▾ 0 more goals This service implements standards discussed here [here](#) and [here](#). Source of the [deployment](#) and the [application](#)

3 from 0 to 3 days ago [todays](#) [last week](#) [future](#) [past](#) [GCNs](#) [arXiv](#) [ATel](#)

Search for...

URI	Title	Facts	Workflows ⓘ
paper:gcn29707 2021-03-24T11:24:43	GRB 210324A: Fermi GBM Final Real-time Localization	DATE NUMBER SUBJECT gcn_authors gcn_from_email gcn_from_name grb_isot instrument mentions_grb mentions_named_grb source	Explore paper origin fetch-workflow-paper:gcn29707 Run INTEGRAL follow-up Run my affinity oda:me oda:cares_for paper:gcn29707 Care Uncare
paper:gcn29715 2021-03-24T19:35:15	GRB 210324B: Fermi GBM Final Real-time Localization	DATE NUMBER SUBJECT gcn_authors gcn_from_email gcn_from_name grb_isot instrument mentions_grb mentions_named_grb source	Explore
paper:gcn29718 2021-03-24T20:10:22	GRB 210324C: Fermi GBM Final Real-time Localization	DATE NUMBER SUBJECT gcn_authors gcn_from_email gcn_from_name grb_isot instrument mentions_grb mentions_named_grb source	Explore
paper:gcn29709	GRB 210324A: Fermi GBM detection		Explore

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.

The screenshot displays the European Open Science Cloud (EOSC) interface. At the top, the EOSC logo and navigation bar are visible. The main content area shows the details for the 'Astronomical Online Data Analysis (AstroODA)' resource, including the CDCI logo, organization information (University of Geneva), and a star rating. Below this, there is a 'swagger' section with a 'default (v27api-doc)' link. The Swagger UI shows the API documentation for 'oda-sdss', including a search bar, navigation tabs (Overview, Collaboration, Files, Datasets, Environments, Settings), and a file browser showing a workflow diagram for 'figure_sky.png'.

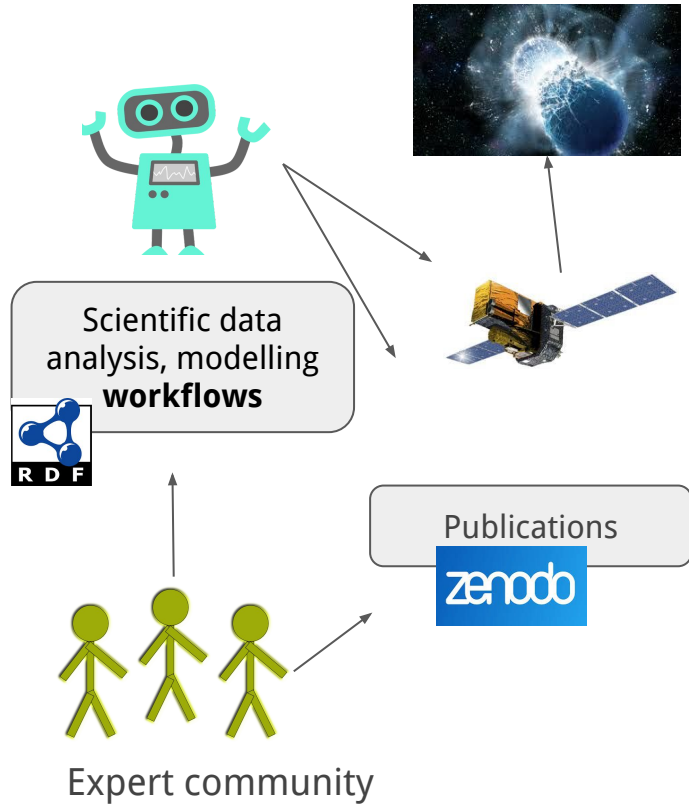
Summary

- 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 offer **adapters for data formats**
- Deriving **data, result provenance**
 - explain data
 - trace history
 - credit and blame creators



*Provenance
RDF PROV-O*

