

COCC EuroScienceGateway

EuroScienceGateway

Volodymyr Savchenko

SKA days 2023

Zurich September 6 2023

EuroScienceGateway: the project

- 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



https://galaxyproject.org/projects/esg/

Galaxy Web-based Analysis Platform ("The Gateway")

Grew out of needs of bioinformatics needs, but reached broad user community: Life Sciences, Materials Science, Climate/Earth science.

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







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 took example of HDF5 to extend Galaxy with **FITS** format: **identification**, **parsing** (astropy), and **visualization** (AladinLite).



Most of the work by Francois Morier-Genoud

https://galaxyproject.org/news/2023-06-20-esg-wp5-astronomy-fits/

Adding IVOA archives, Rucio interface

Galaxy supports S3, webdav, pyfilesystem, etc. OakRidge team added support of Rucio (PASC talk)

We added first interface to query **IVOA TAP archives** from Galaxy (<u>demo video</u>). Integration with ESAP is explored too.

& Astronomical Archives (IVOA) queries astronomical archives through Virtu	al Observatory protocols (Galaxy Version 0.9.0)											
Tool Parameters												
Archive Selection												
Query specific IVOA TAP archive	🖬 Galaxy		🛠 Wor	kflow Visuali	ze Shared Data •	Admin Help	User• 📢 🗃 🏢			Using 4.2 GB		
Archive *	Tools 🌣 -								History	+ = -		
The VO @ ASTRON TAP service	search tools 🛛 🛠 🗙		RESOURCE	S PREV	VIEW AF	CHIVE	https://vo.astron.nl	tap	search datasets	× ×		
Query selection	± Upload Data Get Data	ADQL query : S	ELECT TOP 100 * I	ROM ivoa	obscore	obs id	obs title	A	Unnamed history	1		
ADQL Query Selection	Astronomical Archives (IVOA) queries astronomical archives through Virtual	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B001_0	Archit	ecture		RabbitMQ	
IVOA obscore table query builder	Observatory protocols EBI SRA ENA SRA	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B001_1	i		/	(
Observation target name - optional	EuPathDB server	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B001_2	i)	Cloud storage	
Typically name of the astronomical object observed	HbVar Human Hemoglobin Variants and Thalassemias InterMine server	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B001_3		Galaxy			PULSAR
	modENCODE fly server modENCODE modMine server	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B008_0	i Chobeng	O			
Stop time in MJD - optional	modENCODE worm server MouseMine server	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B008_1			J	d- openo	
	NCBI Datasets Genomes import data from the NCBI Datasets Genomes page	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B008_2	·				
Data product type input choice	Ratmine server	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B008_3	Provider	3		(Rucio Plugin)	
From list	UCSC Archaea table browser	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B015_0	Service		NFS, H	TTP	S3
Data product type *	Upload File from your computer	cube	dirty beam	3	apertif-dr1	190807041	190807041_AP_B015_1	j	00 °D			d ⁺ openD
Image	WormBase server										(
Logical data product type (link to reference in Help section) Collection -senses												Local Storage
Name of the observation collection Observation title (archive specific) - optional								OAK RIDGE	Expe	eriment	Ĺ	Compute Resource
1ost of the work by <i>Fr</i>	ancois Morier-	Genou	ıd									

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:**

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.



Conclusions

- **EuroScienceGateway** is enabling portal/platform access to **OBC** compute and storage assets, involving new communities and adapting to them.
- Astronomy community has already developed and is developing a variety of solutions for FAIR data and workflows, which is why we choose to integrate existing solutions (IVOA, AladinLite, astropy, astroquery, workflow catalogs from SDSC, AstroORDAS, etc) with Galaxy.
- **Galaxy** is a very thoroughly developed Science Platform with exceptional experience, and it is **highly beneficial to learn from it**.
- **EuroScienceGateway** project will help (among many other goals) to see future potential of **Galaxy** platform in Astronomy. It is very likely that owing to its adaptability it can be reasonably **integrated with other solutions in astronomy** on multiple levels: compute, workflow/data catalogs/registries, visualizations, **enriching web-based scientific analysis ecosystem** (see also a <u>note</u>)