EuroScienceGateway

Volodymyr Savchenko

SKA days 2023

Zurich
September 6 2023
EuroScienceGateway: the project

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/

Andrii Neronov, Francois Morier-Genoud, VS
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**

**Interactive tools**

https://usegalaxy.eu/  https://galaxyproject.org/
**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 François Morier-Genoud

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 (demo video). Integration with ESAP is explored too.

Most of the work by Francois Morier-Genoud
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: 🌱WorkflowHub

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 EOSC 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 note)