Obligatory organigram
Expected results

Pulsar Network:
A mature and tested (TRL-9) distributed compute network with demonstrated usage across at least 12 European partners.

ScienceGateways:
6 national Galaxy instances operational and proven by the scientific community with more than 100,000 users.

Communities:
Science Gateway with a custom set of tools and workflows for the Biodiversity/Climate, Materials Science and Astrophysics community.

FAIR data:
EOSC-catalogued FAIR data and workflows that can be found, consumed, created and published (RO-Crate) by users of the EuroScienceGateway, demonstrated for each of the 3 domain-specific use-cases.
What is Galaxy!
Communities

- Material Science
  https://materials.usegalaxy.eu
- Astrophysics
- Biodiversity
  https://assembly.usegalaxy.eu
  https://ecology.usegalaxy.eu
Galaxy as large scale compute infrastructure

- Multi-user environment (77,000 - 2023/10)
  - Caching of reference data
  - Caching of jobs
  - Multi-tiered scheduling (workflows, “users” (limits), jobs) → EuroScienceGateway
- 3,000 tools (Assembly, Jupyter, Alphafold, ...) - a system of systems - a VRE of VREs
- ~2,000,000 container runs a month
- 2.5 TB uploads per day

Open Infrastructure

https://github.com/usegalaxy-eu
On-boarding: APIs, standards, logins

- Users login via LDAP, AD, PAM, CAS, OpenID, ORCID, LS-LOGIN
- Findability / Interoperability
Mirco-services to handle requests

- Users interacting with the a) web frontend or b) a RESTful API
- Galaxy is a ASGI application
Micro-services to handle requests

- Users interacting with the a) web frontend or b) a RESTful API
- Galaxy is an ASGI application
Common theme: It’s actually a bit more complicated ;)

- Resilient and scalable services
Galaxy as large scale compute infrastructure

- Jobs can be submitted to “destinations”
- All common and not so common DRMs
- dependency resolution
Galaxy as large scale compute infrastructure

- Running on all resources
Galaxy as large scale compute infrastructure

- Bridging to all resources at once with Pulsar
- ARC and DIRAC integration in the works
- Meta-Scheduler → EuroScienceGateway
Things that I left out ... time and so ... 

- Uploads and any kind of IO is hard

- Monitoring is essential - it helps with user support, it helps your DevOPS, it helps everyone even your funders ;) [https://stats.galaxyproject.eu](https://stats.galaxyproject.eu)
Things that I left out … time and so …

- STORAGE and DATA
- Distributed storage
- Local storage
- User owned storage
Open Infrastructure
https://github.com/usegalaxy-eu
https://eurosciencegateway.eu
Open Infrastructure
https://github.com/usegalaxy-eu
https://eurosciencegateway.eu

PULSAR
Infrastructure

- 6 national Galaxy instances
- 12 compute sites as part of the decentralised Pulsar network
- Smart meta-scheduler (optimize for performance, carbon footprint, cost ...)
  - Integrate with ARC and DIRAC
- BYOC - spin up your own Pulsar, ARC endpoint in your academic or commercial cloud - let Galaxy use it
- BYOD - user based object stores (you have an S3 bucket from your University, put it into Galaxy)
usegalaxy.* community expanding
Data
Data

- RO- Crate into Galaxy and out
- Workflows from WorkflowHub into Galaxy
- Workflows from Galaxy to WorkflowHub
- Galaxy artefacts into the OpenAIRE KnowledgeGraph
Communities sharing one coherent framework

- rna.usegalaxy.eu
- clipseq.usegalaxy.eu
- metagenomics.usegalaxy.eu
- hicexplorer.usegalaxy.eu
- cheminformatics.usegalaxy.eu
- proteomics.usegalaxy.eu
- imaging.usegalaxy.eu
- metabolomics.usegalaxy.eu
- ecology.usegalaxy.eu
- nanopore.usegalaxy.eu
- singlecellomics.usegalaxy.eu
- humancellatlas.usegalaxy.eu
- virology.usegalaxy.eu
- climate.usegalaxy.eu
- streetscience.usegalaxy.eu
- ml.usegalaxy.eu
BY-COVID

Infectious Diseases Toolkit

Showcase

Pathogen characterisation
Socioeconomic data
Human biomolecular data
Human clinical and health data

Showcase

An automated SARS-CoV-2 genome surveillance system built around Galaxy

An automated, modular system for large-scale FAIR analysis of SARS-CoV-2 sequencing data analysis powered by the Galaxy platform.

Affiliations: 🇩🇪 🇪🇺
### Biodiversity and the BGE project

**Project Highlight:**

The Vertebrate Genomes Project in Galaxy

Generating near error-free genome assemblies for more than 70,000 vertebrate species!
Even more diversity

even more sharing, more empowering

- Build a coherent analysis platform for all researchers
- More citizen science projects
- Outreach is hard - and we are not outreach experts
- How do we scale user-support to 100k users?
Stronger together: FAIR-EASE and EuroScienceGateway join forces

06 February 2023

Two Horizon Europe projects directed toward the development and deployment of the European Open Science Cloud (EOSC)—FAIR-EASE and EuroScienceGateway (ESG)—have agreed to closely cooperate over the next three years in order to extend the open source Galaxy platform for FAIR data analysis to a broader group of scientific user communities.

The cooperation will include shared compute resources, joint workshops, and a merging of tools and workflows from FAIR-EASE to be run via the distributed computing network being developed by the ESG project. The aim is to evaluate if FAIR-EASE’s Earth Analytical Lab (EAL) can be based on the ESG.

“Synchronising the work of ESG and FAIR-EASE is good news for the users of EOSC,” said Björn Grüning, project coordinator of ESG. “We believe that if ESG and FAIR-EASE combine our efforts we can achieve more for our users and utilise European resources more effectively. This is a way of working that the European Commission and the EOSC Association have been encouraging among the EOSC-related projects.”
Let's discuss VREs

- **Galaxy as a VRE**
  - Galaxy can and is deployed as VRE for single user

- **Galaxy is a system of systems**
  - We offer 3000 “VRE” (we call it tools) on the European Galaxy instance
    - Data importer, Data exporter
    - Traditional HPC tools
    - Text Processing tools
    - Jupyter Notebooks, RStudio

- **Galaxy can schedule all kinds of “VRE” (like Jupyter)**
Galaxy Interactive Tools (GxITs): A Wormhole Between the Worlds

“HPC” / CLI tools
pre-built tools, rigid interface, Galactic scale

Interactive tools
Complete freedom! Dynamic outputs, no rules, less scale
GxITs: Use RStudio & Jupyter directly within Galaxy
live.usegalaxy.eu
Interactive Tools are just tools that start containers.

Galaxy knows how to direct user requests to cluster where the tools are running.
More technical
Interactive tools in COVID-19 variant calling analysis
Climate analysis on usegalaxy.eu

Welcome to the climate science community

The Climate Science workbench is a comprehensive set of simple tools and streamlined workflows. The workflow is based on the Galaxy framework, which guarantees simple access, easy extensibility, robust security vs. personal and facility needs, and streamlined workflows dependent on community guidelines.

The current implementation covers the tools dedicated to different research areas of climate science. More tools are coming soon.

The list of tools is maintained by users.

Content
1. Welcome to the climate science community
   1.1. Using Galaxy
   1.2. Training
   1.3. Interactive tools
   1.4. Climate variables
   1.5. Repository use
   1.6. Interactive learning environment
   1.7. All workshops

Get started
Are you new to Galaxy? Or returning after a long time, and looking for help to get started? Take a guided tour through Galaxy’s own interface.

Training
1. Galaxy 101 for researchers
2. Climate Galaxy Tutorial

Want to learn more about Galaxy? Check out the following hands-on tutorials from the Galaxy Training Network.

We are interested in hearing from you. Do you see any missing or broken links? Help us improve Galaxy!

Available tools

Interactive tools
- Python netCDF viewer
- A tool to view climate data: Interactive tools (Land & Climate: usegalaxy.eu) is an integrated, Python-based tool that supports Python, Jupyter, and Python-based software.
JupyterLab for Ocean / Atmosphere / Land / Climate

Use cases

- Development of climate models
- Co-design for new Galaxy tool development
- Data analysis and visualization
Panoply netCDF viewer

Use cases

- Data visualization
- Image saving
- Time series video creation
Ecology with Interactive Tools: 3 use cases

Inside a complete analysis

Biodiversity and environmental data collection

Selection / Filtering

Analysis via modelling
R Shiny Wallace

Data visualisation

Cartography + visualisation
R Shiny GeoExplorer

Data cleaning

Data cleaning
OpenRefine
GPU-enabled VREs for everyone

An accessible infrastructure for artificial intelligence using a Docker-based JupyterLab in Galaxy

Image segmentation of COVID-19 CT scans

Prediction of 3D structures of proteins
Single-cell data analysis

When offline limitation (RAM, dataset size, etc)