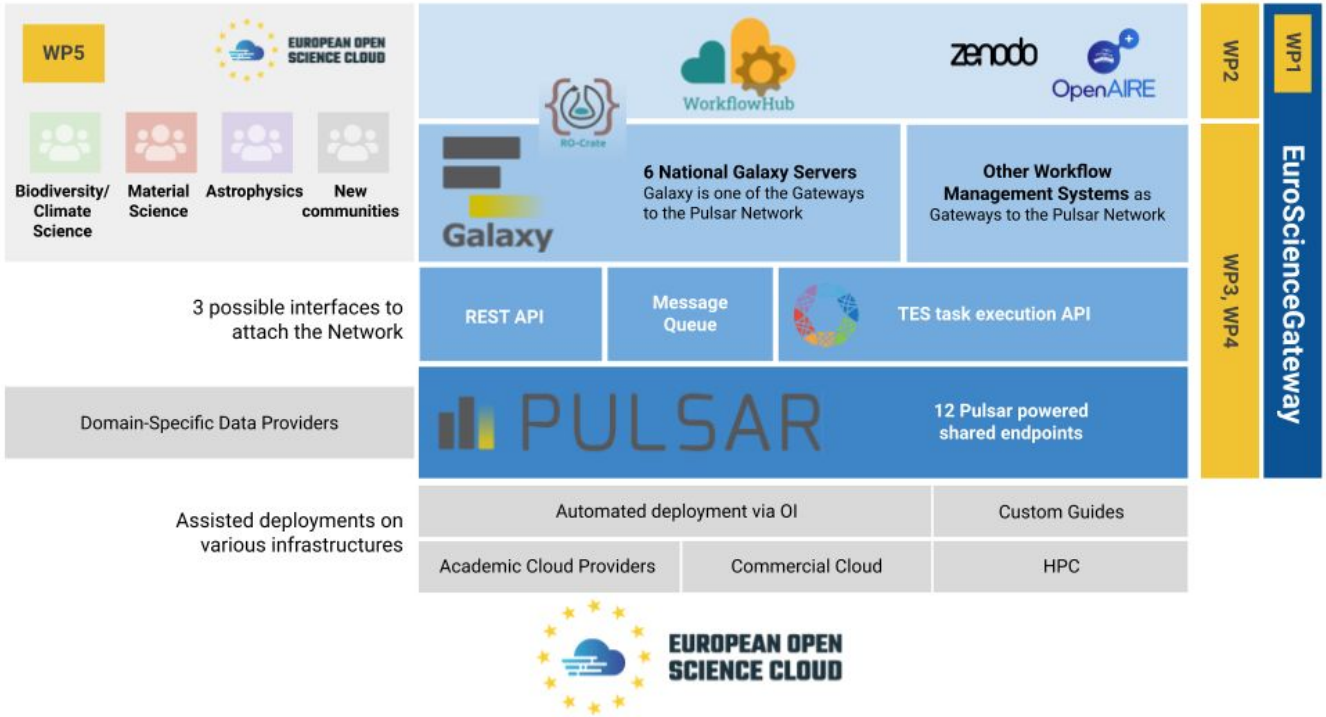
https://eurosciencegateway.eu

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**.

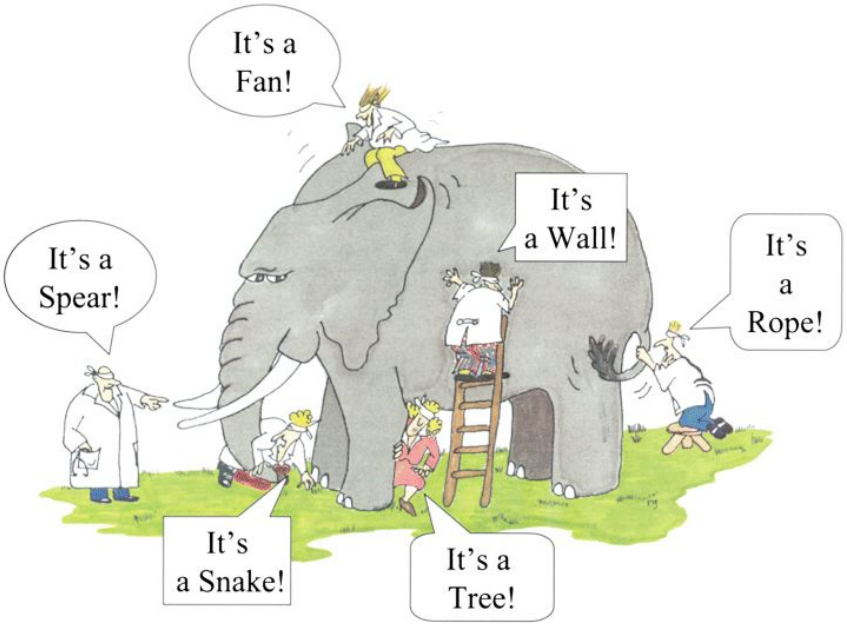
Communities:

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

FAIR data:

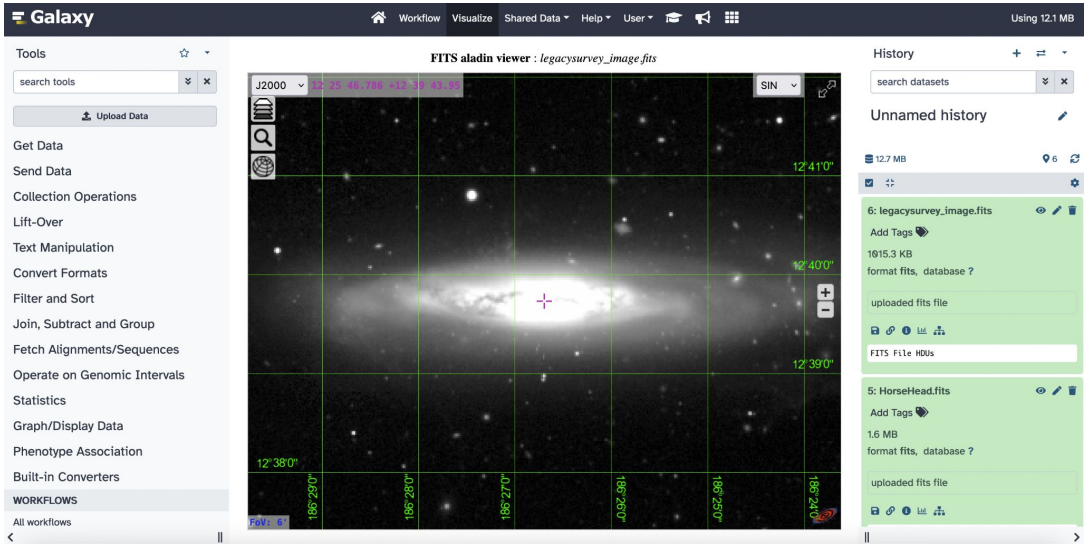
EOSC-catalogued **FAIR data and workflows** that can be found, consumed, created and published (**RO-Crate**) by users of the EuroScienceGateway, demonstrated for each of the 3 domain-specific use-cases.

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

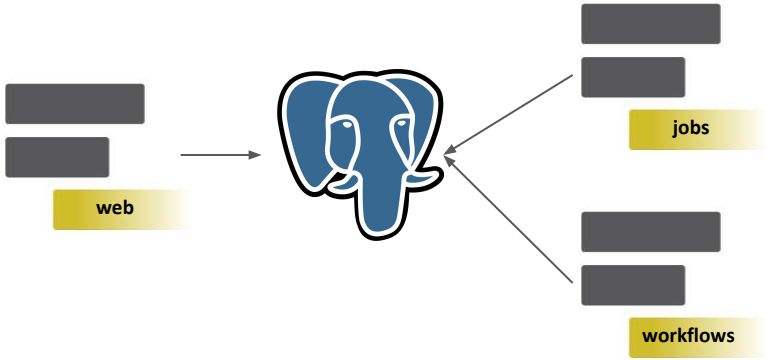
On-boarding: APIs, standards, logins

- Users login via LDAP, AD, PAM, CAS, OpenID, ORCID, LS-LOGIN
- Findability / Interoperability

Global Alliance
for Genomics & Health

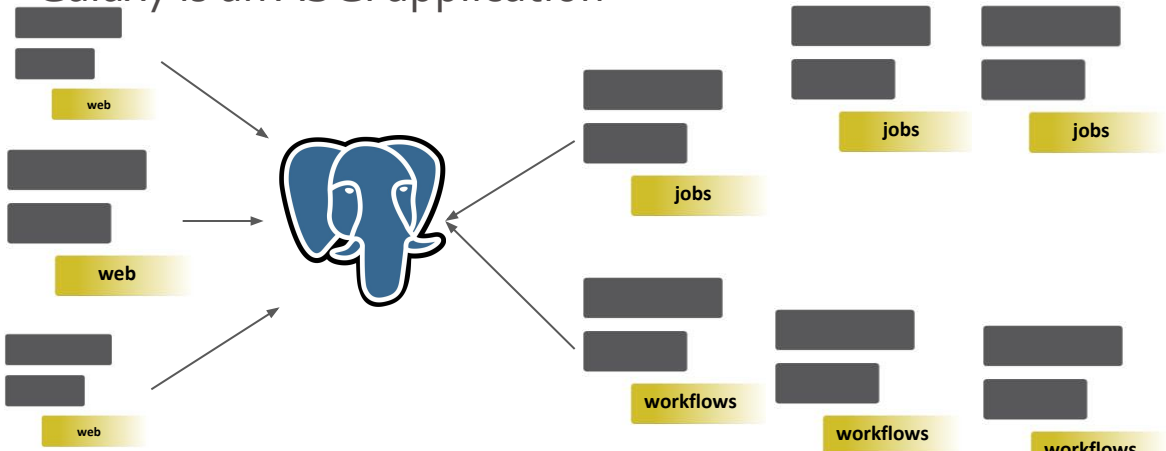
Mirco-services to handle requests

- Users interacting with the a) web frontend or b) a RESTful API
- Galaxy is a ASGI application



Micro-services to handle requests

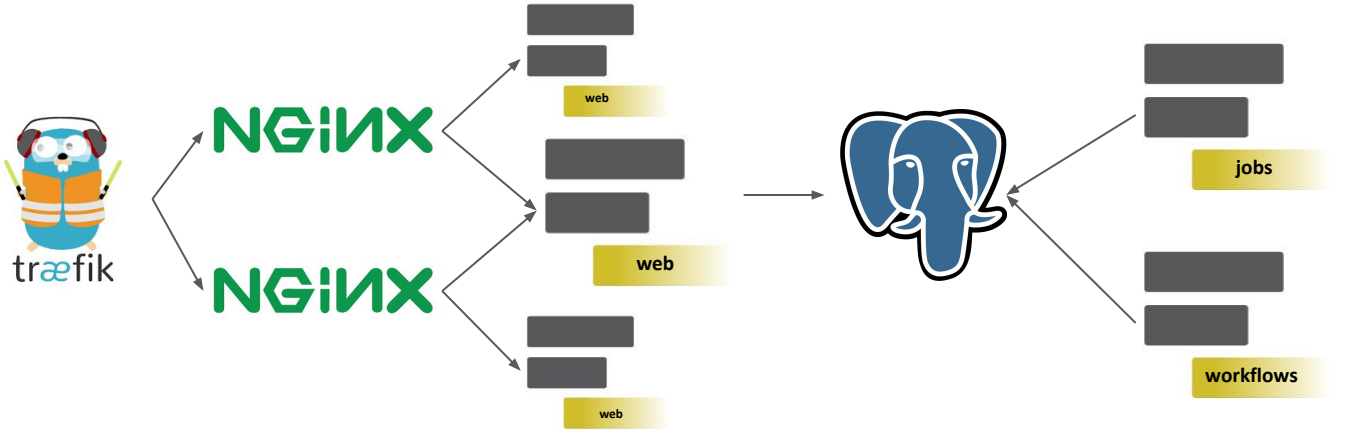
- Users interacting with the a) web frontend or b) a RESTful API
- Galaxy is an ASGI application



OPENAPI Global Alliance for Genomics & Health schema.org LS LOGIN Galaxy Training Network

Common theme: It's actually a bit more complicated ;)

- Resilient and scalable services



OPENAPI Global Alliance for Genomics & Health schema.org LS LOGIN Galaxy Training Network

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

openstack

kubernetes

aws

Google Cloud

Microsoft Azure

docker

HTCondor Software Suite

slurm workload manager

kubernetes

web

jobs

workflows

NGINX

træfik

OPENAPI

Global Alliance for Genomics & Health

schema.org

LS LOGIN

Galaxy Training Network

Galaxy as large scale compute infrastructure

- Bridging to all resources at onces with Pulsar
- ARC and DIRAC integration in the works
- Meta-Scheduler → EuroScienceGateway

HPC openstack kubernetes aws Google Cloud Microsoft Azure

docker HTCondor Software Suite slurm workload manager kubernetes

web jobs workflows NGINX træfik

OPENAPI Global Alliance for Genomics & Health schema.org LS LOGIN Galaxy Training Network

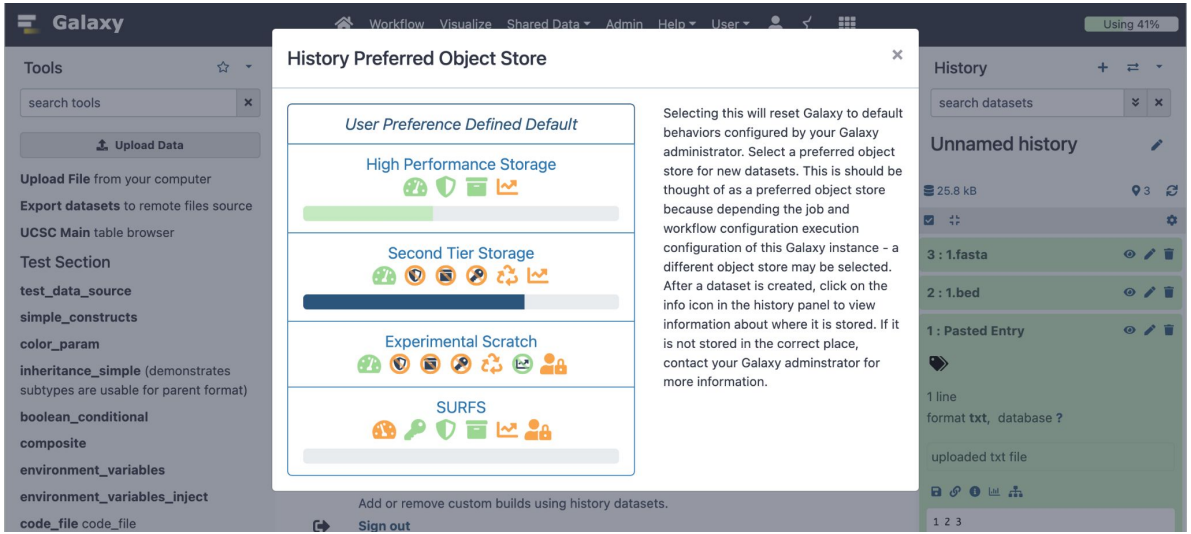
Things that I left out … time and so …

- 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

Things that I left out … time and so …

- STORAGE and DATA
- Distributed storage
- Local storage
- User owned storage



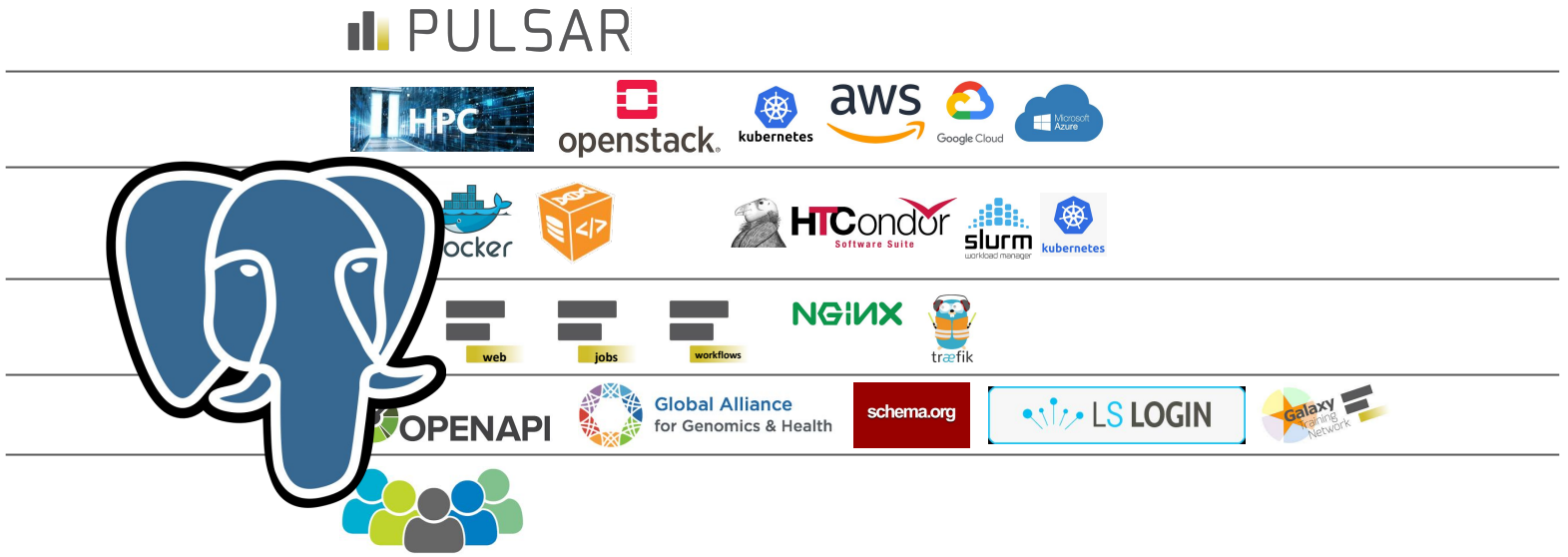
Amazon S3

iRODS

10
01

Open Infrastructure
https://github.com/usegalaxy-eu
https://eurosciencegateway.eu

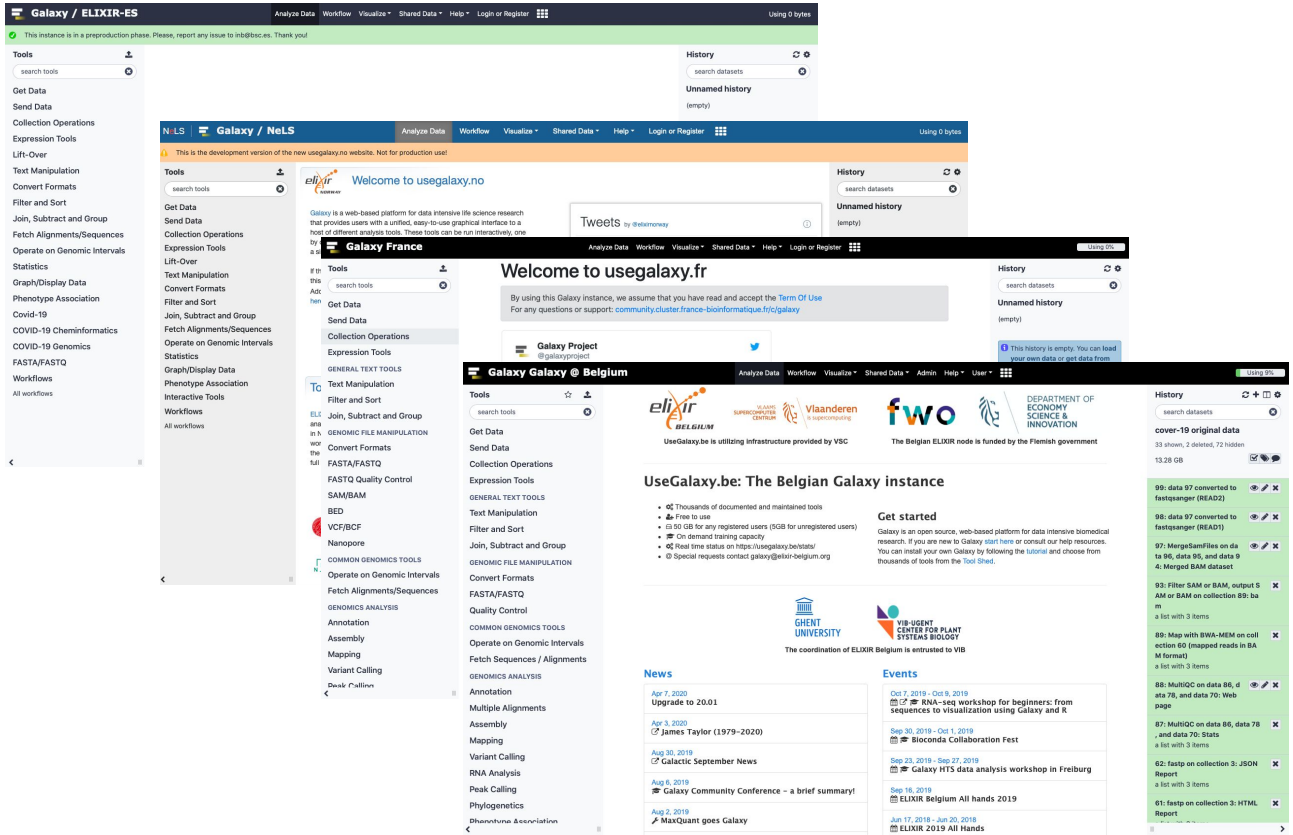
Open Infrastructure
https://github.com/usegalaxy-eu
https://eurosciencegateway.eu



Infrastructure

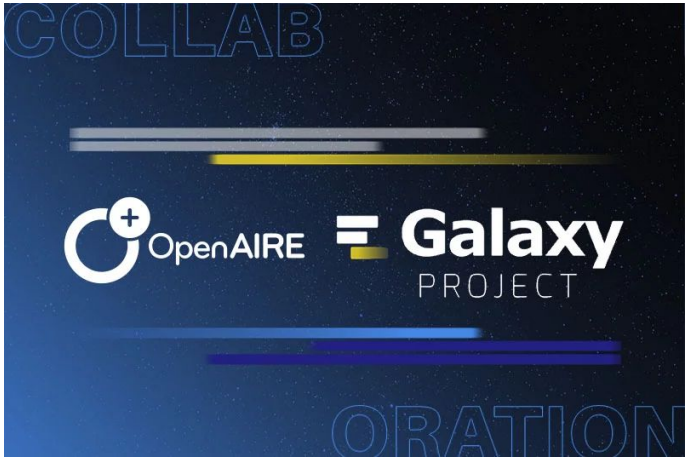
- 6 national Galaxy instances
- 12 compute sites as part of the decentralised Pulsar network
- Smart meta-scheduler (optimize for performance, carbon footprint, cost ...)
  - Integrate with ARC and DIRAC
- BYOC - spin up your own Pulsar, ARC endpoint in your academic or commercial cloud - let Galaxy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

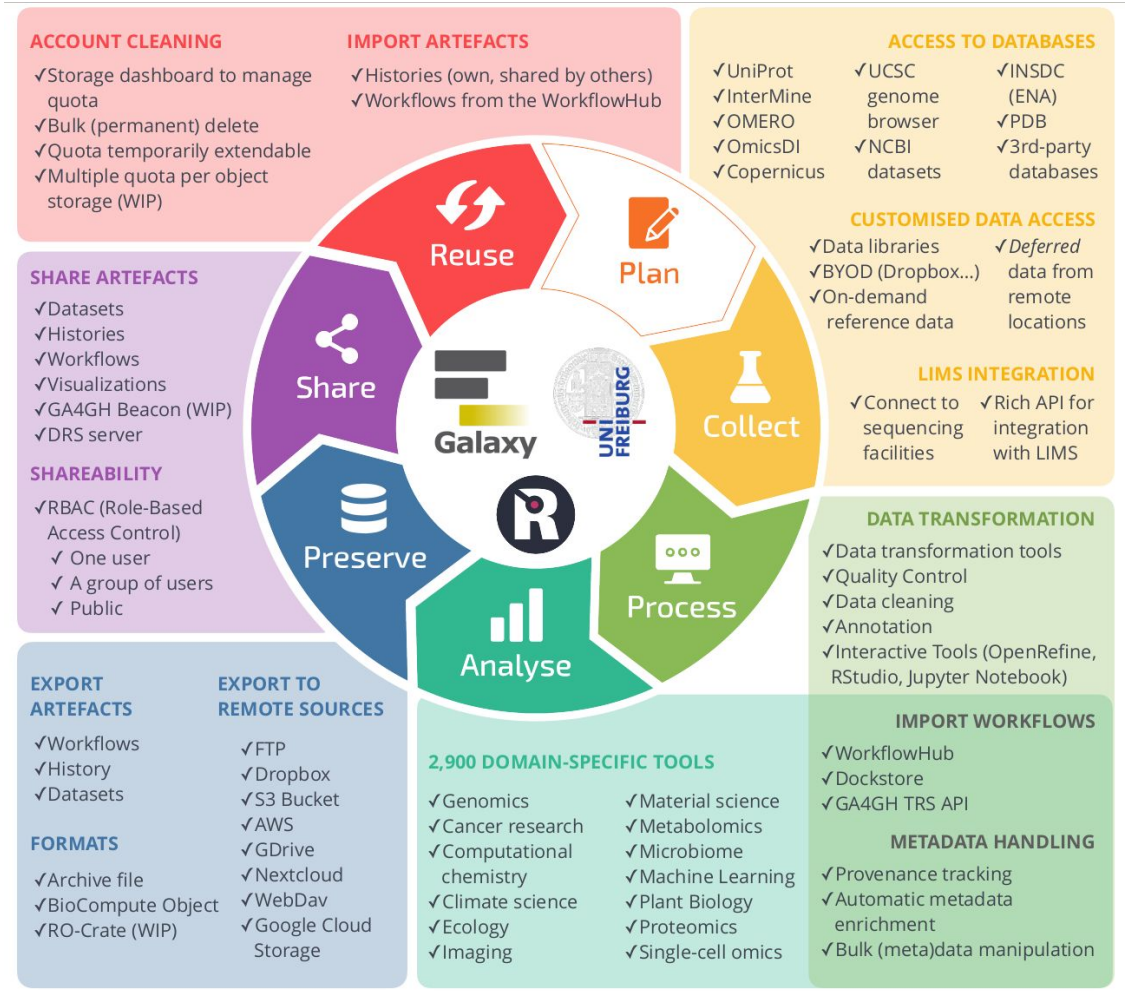
usegalaxy.* community expanding



usegalaxy.es
usegalaxy.no
usegalaxy.fr
usegalaxy.be
usegalaxy.ee
usegalaxy.it
usegalaxy.eu
usegalaxy.cz
usegalaxy.org
usegalaxy.co.au

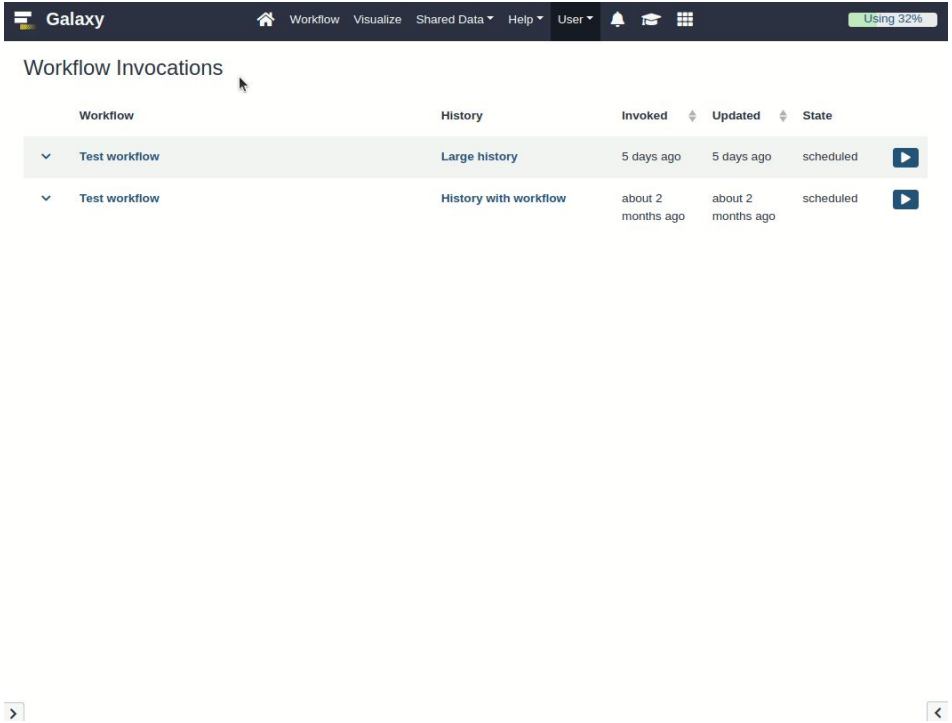
Data


COLLAB
OpenAIRE
Galaxy
PROJECT
ORATION


ACCOUNT CLEANING
✓Storage dashboard to manage quota
✓Bulk (permanent) delete
✓Quota temporarily extendable
✓Multiple quota per object storage (WIP)
IMPORT ARTEFACTS
✓Histories (own, shared by others)
✓Workflows from the WorkflowHub
ACCESS TO DATABASES
✓UniProt
✓InterMine
✓OMERO
✓OmicsDI
✓Copernicus
✓UCSC genome browser
✓NCBI datasets
✓INSDC (ENA)
✓PDB
✓3rd-party databases
CUSTOMISED DATA ACCESS
✓Data libraries
✓BYOD (Dropbox...)
✓On-demand reference data
✓Deferred data from remote locations
LIMS INTEGRATION
✓Connect to sequencing facilities
✓Rich API for integration with LIMS
SHARE ARTEFACTS
✓Datasets
✓Histories
✓Workflows
✓Visualizations
✓GA4GH Beacon (WIP)
✓DRS server
SHAREABILITY
✓RBAC (Role-Based Access Control)
✓ One user
✓ A group of users
✓ Public
Reuse
Plan
Share
Collect
Galaxy
UNI FREIBURG
Preserve
Process
Analyse
DATA TRANSFORMATION
✓Data transformation tools
✓Quality Control
✓Data cleaning
✓Annotation
✓Interactive Tools (OpenRefine, RStudio, Jupyter Notebook)
EXPORT ARTEFACTS
✓Workflows
✓History
✓Datasets
FORMATS
✓Archive file
✓BioCompute Object
✓RO-Crate (WIP)
EXPORT TO REMOTE SOURCES
✓FTP
✓Dropbox
✓S3 Bucket
✓AWS
✓GDrive
✓Nextcloud
✓WebDav
✓Google Cloud Storage
2,900 DOMAIN-SPECIFIC TOOLS
✓Genomics
✓Cancer research
✓Computational chemistry
✓Climate science
✓Ecology
✓Imaging
✓Material science
✓Metabolomics
✓Microbiome
✓Machine Learning
✓Plant Biology
✓Proteomics
✓Single-cell omics
IMPORT WORKFLOWS
✓WorkflowHub
✓Dockstore
✓GA4GH TRS API
METADATA HANDLING
✓Provenance tracking
✓Automatic metadata enrichment
✓Bulk (meta)data manipulation

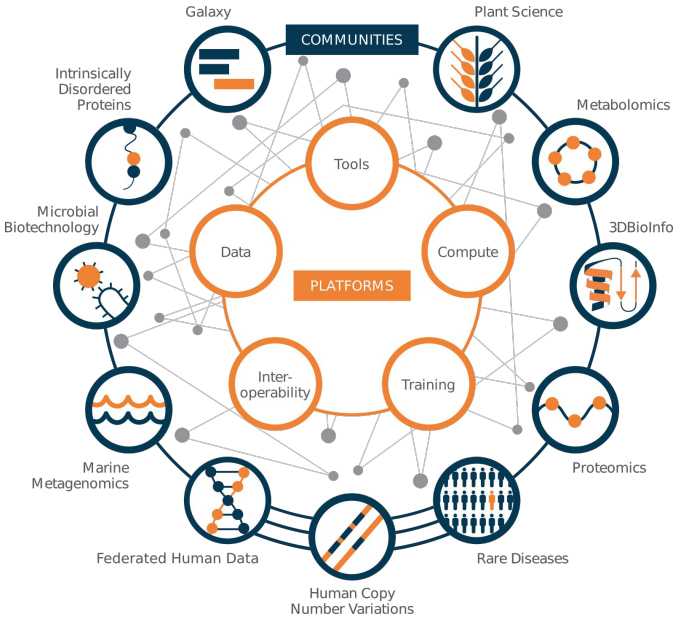
Data

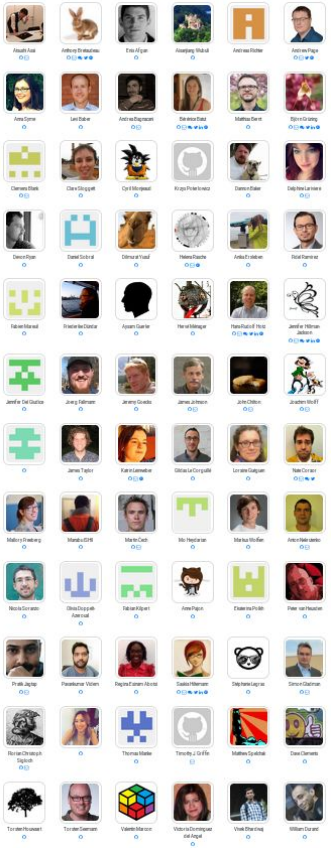
- RO- Crate into Galaxy and out
- Workflows from WorkflowHub into Galaxy
- Workflows from Galaxy to WorkflowHub
- Galaxy artefacts into the OpenAIRE KnowlegdeGraph



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



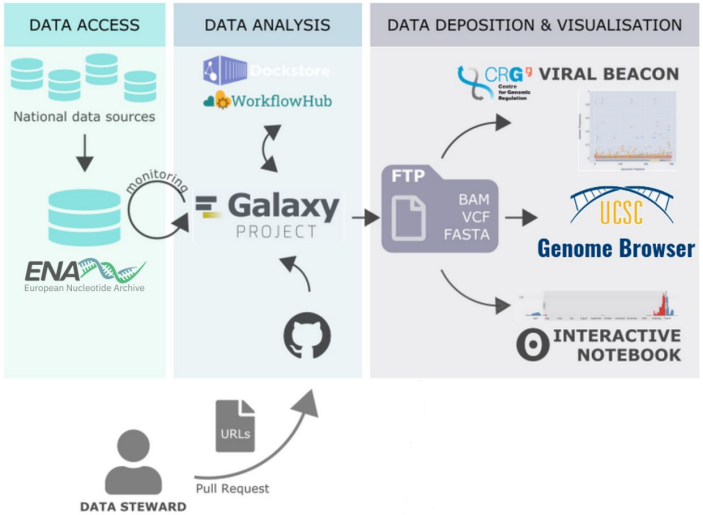


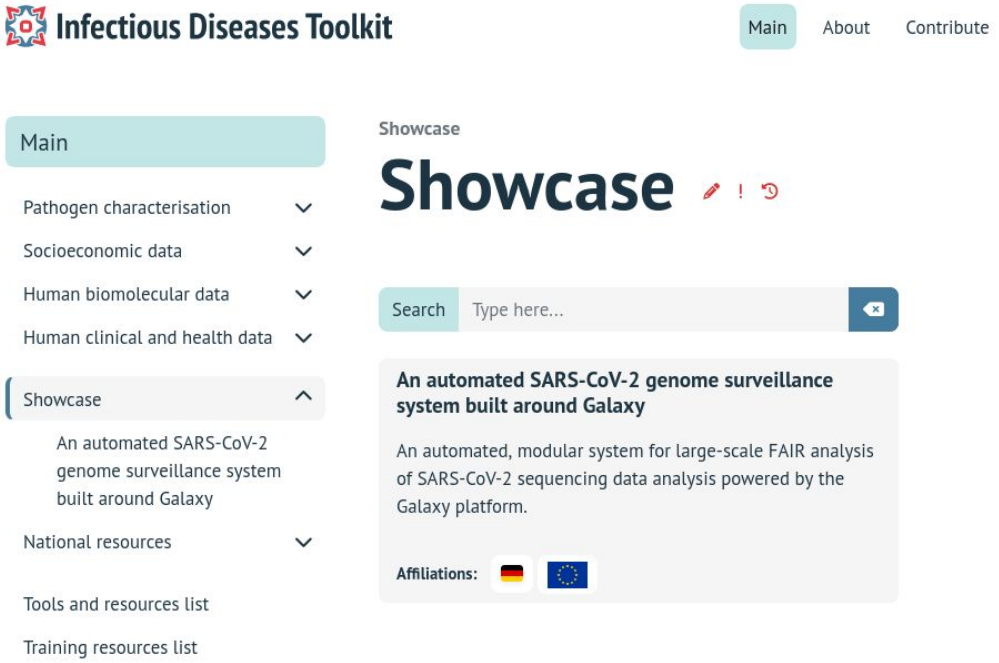
HUMAN CELL ATLAS

ERGA
EUROPEAN REFERENCE GENOME ATLAS

BY-COVID

ht[illegible]axy





Biodiversity and the BGE project

BGE, VGP, ERGA ❤️

Class	Order	Species	Size (Mb)	Contig NG50 (Mb)	Scaffold NG50 (Mb)	Gaps (Mb)	Het (%)	Repeat (%)	Merqury	Busco (%)
Mammalia	Sirenia	*Dugong dugon*	3,159	58	138	0	0.17	34.8	99.70	97.80
		Dugong dugon	3,159	58	177	0	0.17	34.8	99.70	97.90
	Proboscidea	*Loxodonta africana*	3,396	0	0	0	0.32	26.8	99.34	96.50
		Loxodonta africana	3,396	0	0	0	0.32	26.8	99.34	97.80
	Pholidota	*Manis pentadactyla*	2,729	31	139	15	0.21	32.2	99.44	98.00
		Manis pentadactyla	2,729	29	132	31	0.21	32.2	99.44	96.60
	Eulipotyphla	*Sorex araneus*	2,231	4	338	59	1.14	27.0	99.26	96.30
	Diprotodontia	*Macropus eugenii*	3,313	16	488	10	0.33	25.0	0.00	98.00
	Didelphimorphia	*Monodelphis domestica*	3,217	5	519	33	0.81	30.6	99.44	96.50
	Dermoptera	*Cynocephalus volans*	2,715	86	125	0	0.61	26.4	99.54	98.20
	Cingulata	*Dasypus novemcinctus*	3,235	16	115	30	0.60	32.5	99.36	96.40
		Dasypus novemcinctus	3,235	15	90	74	0.60	32.5	99.36	93.00
	Carnivora	*Neofelis nebulosa*	2,396	54	149	1	0.36	19.9	99.65	97.50
	Artiodactyla	*Mesoplodon densirostris*	3,125	45	79	1	0.29	42.4	99.24	97.40
		Hippopotamus amphibius	2,553	72	168	0	0.00	23.0	99.60	98.20
		Hippopotamus amphibius	2,553	78	184	0	0.00	23.0	99.60	98.20
		Eubalaena glacialis	2,662	39	87	5	0.23	30.4	99.60	96.00
		Eubalaena glacialis	2,662	33	84	49	0.22	30.4	99.60	97.60
		Eschrichtius robustus	2,729	48	109	0	0.06	32.4	99.60	97.80
		Balaenoptera edeni	2,701	36	108	15	0.06	30.3	99.69	97.80
Aves	Sphenisciformes	*Spheniscus humboldti*	1,247	16	88	1	0.43	12.6	99.63	96.00
	Rheiformes	*Rhea pennata*	1,255	19	80	1	0.45	11.5	99.35	96.90
	Psittaciformes	*Ara ararauna*	1,217	0	0	0	0.32	16.1	99.71	97.30
		Ara ararauna	1,217	0	0	0	0.32	16.1	99.71	96.80
	Passeriformes	*Taeniopygia guttata*	1,010	10	71	5	1.47	10.8	0.00	0.00
		Poecile atricapillus	1,049	8	117	14	0.89	10.5	99.48	0.00
		Poecile atricapillus	1,049	10	76	1	0.89	10.5	99.48	0.00
		Melospiza georgiana	1,335	8	32	14	1.17	10.4	98.43	96.00
		Dixiphia pipra	990	11	76	0	1.16	6.8	99.55	93.10
		Dixiphia pipra	990	14	76	0	1.16	6.8	99.55	95.20
		Corvus hawaiiensis	1,082	23	69	17	0.34	11.0	0.00	95.30
		Ammodramus nelsoni	1,191	12	40	5	0.51	21.3	99.26	96.80
		Ammodramus caudacutus	1,225	8	40	4	0.24	21.4		96.10
		Acridotheres tristis	1,354	7	30	3	0.56	30.0	98.48	95.60
	Opisthocomiformes	*Opisthocomus hoazin*	1,317	14	88	0	0.58	19.9	99.54	96.20
		Opisthocomus hoazin	1,317	17	81	0	0.58	19.9	99.54	91.60
	Gruiformes	*Rhynochetos jubatus*	1,313	11	35	1	0.60	22.5	99.40	96.20
	Coliiformes	*Colius striatus*	1,130	2	5	5	0.65	12.6	99.66	96.20
		Colius striatus	1,130	0	2	46	0.65	12.6	99.66	89.70
	Caprimulgiformes	*Podargus strigoides*	1,093	15	85	2	1.31	12.0	99.76	96.10
Lepidosauria	Squamata	*Vipera latastei*	1,444	46	203	40	0.52	35.0	0.00	0.00
		Liasis olivaceus	1,448	75	204	5	0.19	17.7	99.68	97.80
		Eulepetes europaea	1,590	34	109	6	0.92	27.1	99.43	97.30
		Erythrolamprus reginae	1,908	3	255	0	NaN	38.4	98.41	95.10
		Candoia aspera	1,472	58	209	0	0.49	31.2	99.57	97.90
		Candoia aspera	1,472	52	219	0	0.49	31.2	99.57	97.50
Reptilia	Testudines	*Sternotherus odoratus*	2,287	39	293	58	0.21	22.3	99.65	97.30
		Sternotherus odoratus	2,287	44	256	4	0.21	22.3	99.65	97.90
	Crocodylia	*Emys blandingii*	2,272	81	138	17	0.28	27.6	99.38	98.10
		Gavialis gangeticus	2,301	84	298	1	0.12	22.8	99.70	97.70
		Gavialis gangeticus	2,301	73	256	1	0.12	22.8	99.70	97.90
Amphibia	Anura	*Spea bombifrons*	959	21	109	6	1.73	28.6	99.52	95.80
		Pseudophryne corroboree	8,455	7	793	23	0.47	73.8	97.87	92.50
		Hyla sarda	3,750	5	420	0	1.08	61.2	98.29	92.20
		Hyla sarda	3,750	4	505	0	1.08	61.2	98.29	91.80
		Gastrophryne carolinensis	4,096	5	518	0	1.45	62.0	98.35	93.10
		Gastrophryne carolinensis	4,096	5	533	0	1.45	62.0	98.35	94.00
Actinopteri	Siluriformes	*Trichomycterus rosablanca*	999	0	0	0	0.13	34.1	99.40	95.20
		Trichomycterus rosablanca	999	0	0	0	0.13	34.1	99.40	97.60
	Scombriformes	*Neoarius graeffei*	2,429	15	80	25	0.49	68.0	97.84	97.10
	Perciformes	*Scomber japonicus*	689	7	25	1	1.22	19.9	99.56	97.10
		Lycodopsis pacificus	590	5	29	0	0.86	25.8	99.33	97.60
	Lampriformes	*Lampris incognitus*	1,327	1	61	1	1.06	50.8	98.79	95.80
		Lampris incognitus	1,327	1	67	1	1.06	50.8	98.79	96.10
	Characiformes	*Salminus brasiliensis*	0	6	41	20	0.00	0.0	0.00	97.50
		Salminus brasiliensis	0	6	39	20	0.00	0.0	0.00	97.40
		Hoplias malabaricus	1,007	5	57	61	0.67	33.1	99.25	97.20
		Hoplias malabaricus	1,007	4	57	90	0.67	33.1	99.25	96.40
	Atheriniformes	*Odontesthes bonariensis*	924	10	31	0	0.43	35.7	99.20	98.70
Chondrichthyes	Myliobatiformes	*Mobula japanica*	3,415	10	74	67	0.36	43.7	99.33	91.60
		Mobula japanica	3,415	11	74	35	0.36	43.7	99.33	93.60
		Hypanus sabinus	1,472	5	159	0	0.44	43.5	99.23	90.80
		Hypanus sabinus	1,472	5	153	0	0.44	43.5	99.23	93.40

Project Highlight:

The Vertebrate Genomes Project in Galaxy

VGP VERTEBRATE GENOMES PROJECT

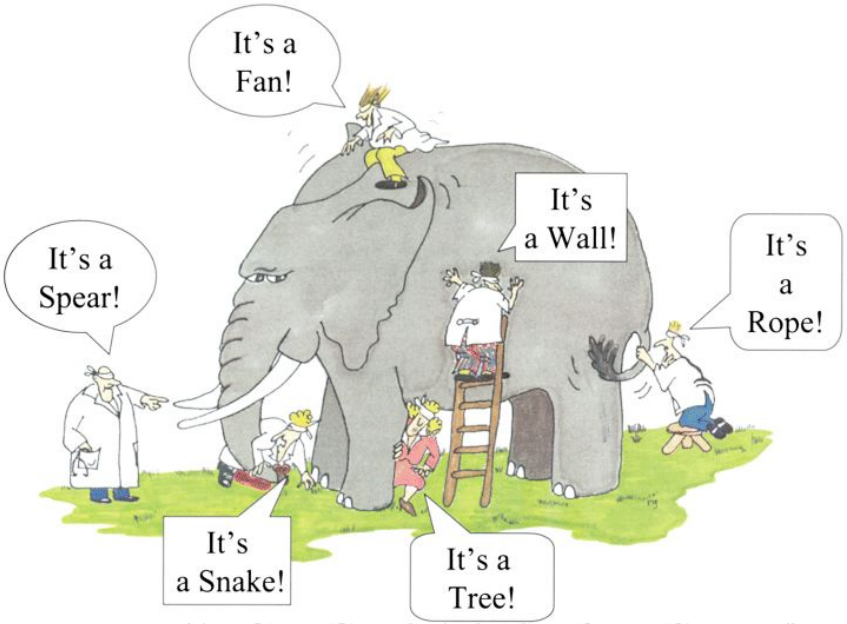
A PROJECT OF THE G10K CONSORTIUM

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.

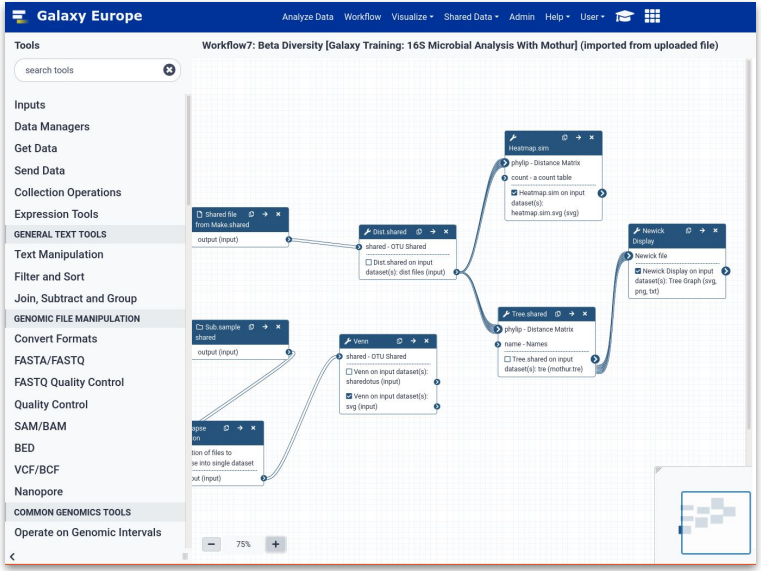
The cooperation will include shared compute resources, joint workshops, and a merging of tools and workflows from FAIR-EASE to be run via the distributed computing network being developed by the ESG project. The aim is to evaluate if FAIR-EASE's Earth Analytical Lab (EAL) can be based on the ESG.

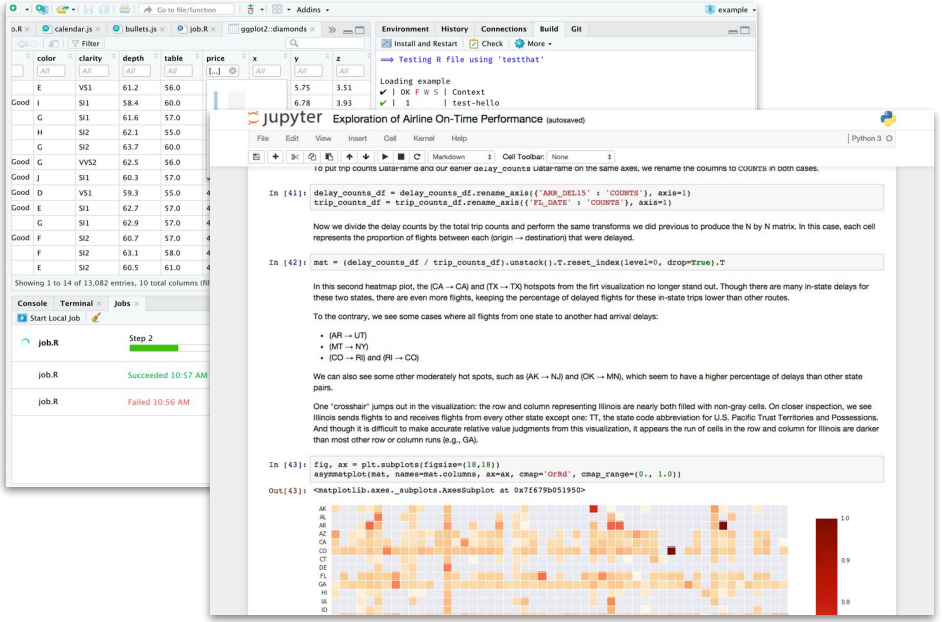
"Synchronising the work of ESG and FAIR-EASE is good news for the users of EOSC," said Björn Grüning, project coordinator of ESG. "We believe that if ESG and FAIR-EASE combine our efforts we can achieve more for our users and utilise European resources more effectively. This is a way of working that 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

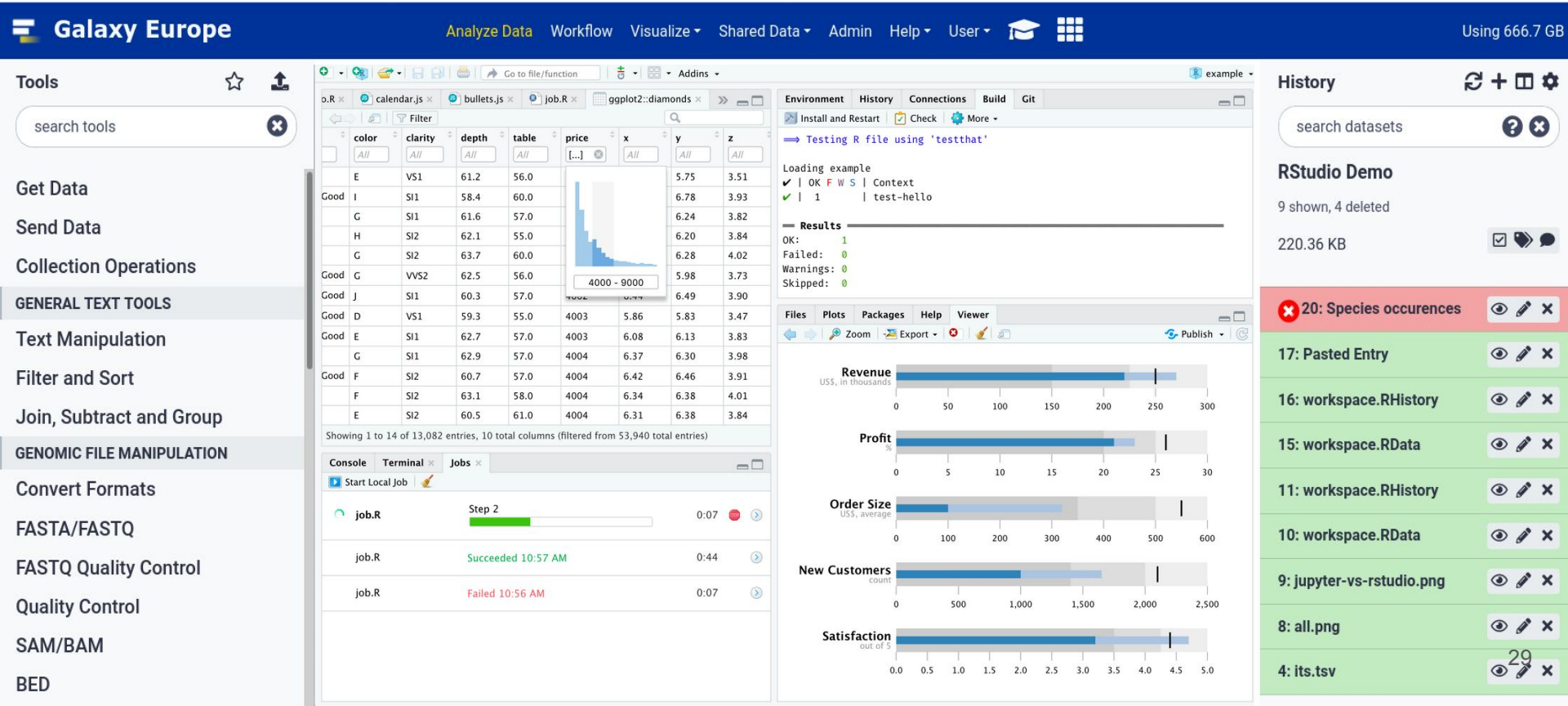




"HPC" / CLI tools
pre-built tools, rigid interface, Galactic scale

Interactive tools
Complete freedom! Dynamic outputs, no rules, less scale

GxITs: Use RStudio & Jupyter directly within Galaxy



Galaxy / Live

Analyze Data Workflow Visualize Shared Data Admin Help User

Using 666.7 GB

Jupyter lab for Python, R and Julia

RStudio with basic R packages

SPARQL query interface

BAM IOBIO

Examine sequence alignment in seconds

live.usegalaxy.eu

VCF IOBIO

Examine your variant file in seconds

Explore single-cell transcriptomics data

Ethercalc

A web spreadsheet

PHINCH

Visualise large biological datasets (microbiomes, metagenomes, etc.)

History

search datasets

RStudio Demo

8 shown, 5 deleted

220.36 KB

17: Pasted Entry

16: workspace.RHistory

15: workspace.RData

11: workspace.RHistory

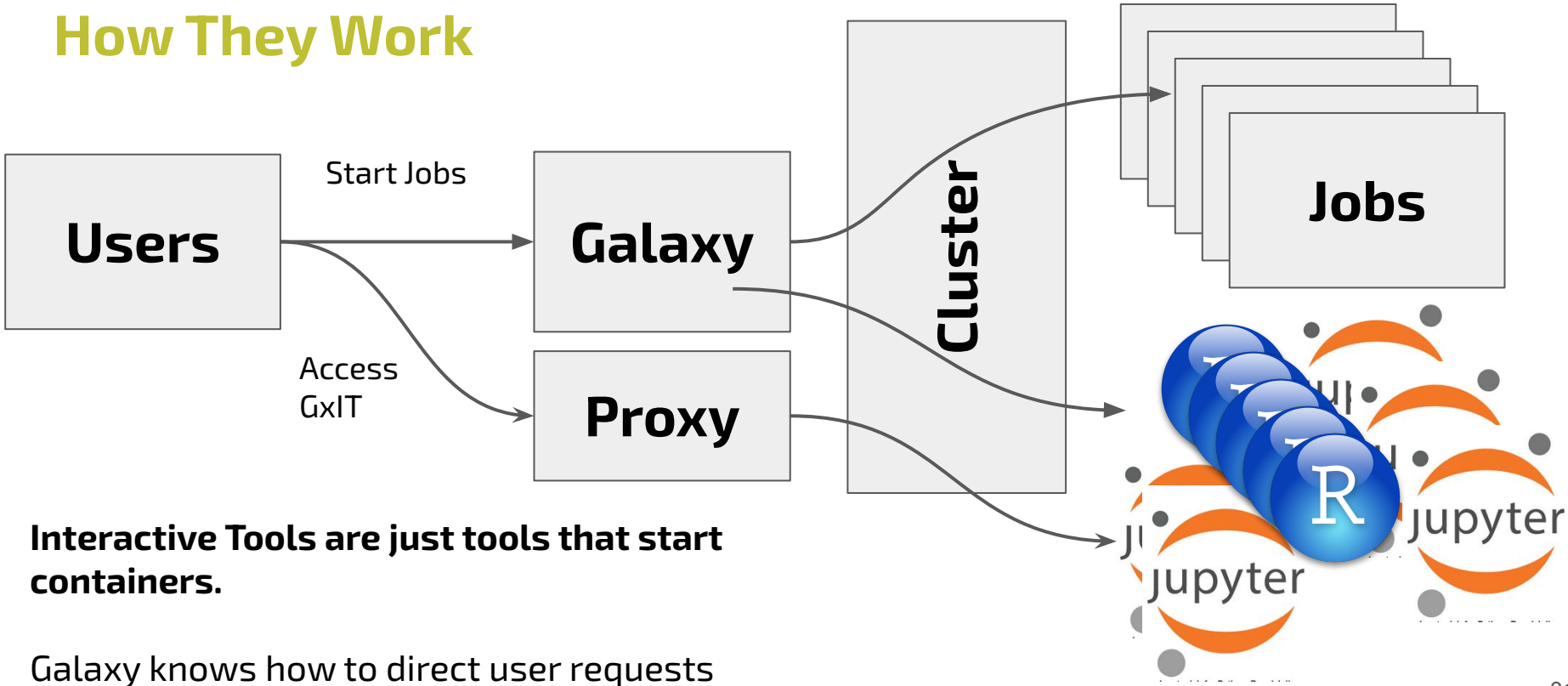
10: workspace.RData

9: jupyter-vs-rstudio.png

8: all.png

4: its.tsv

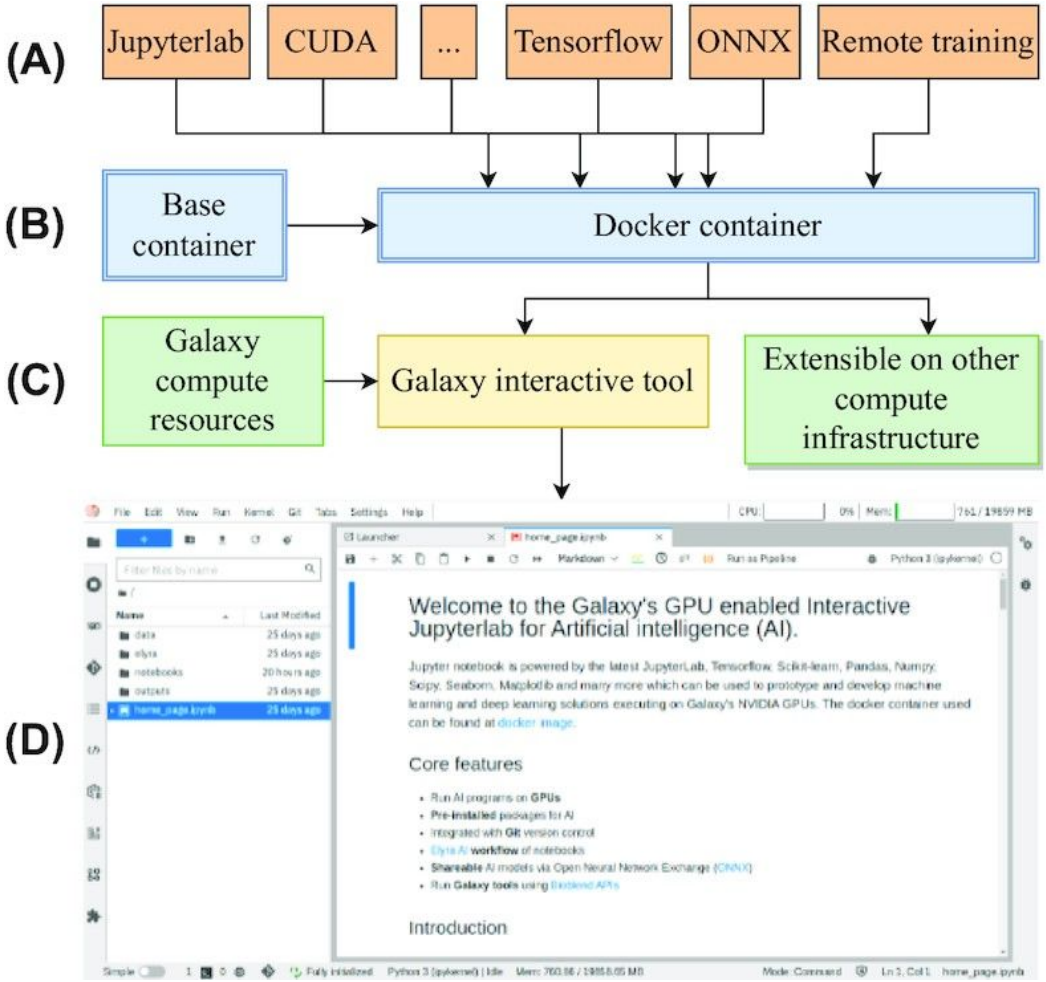
How They Work



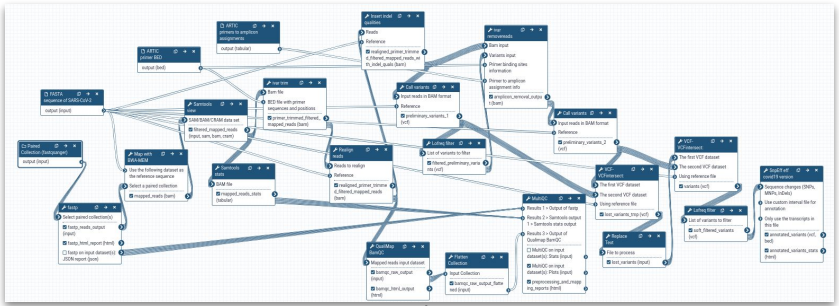
Interactive Tools are just tools that start containers.

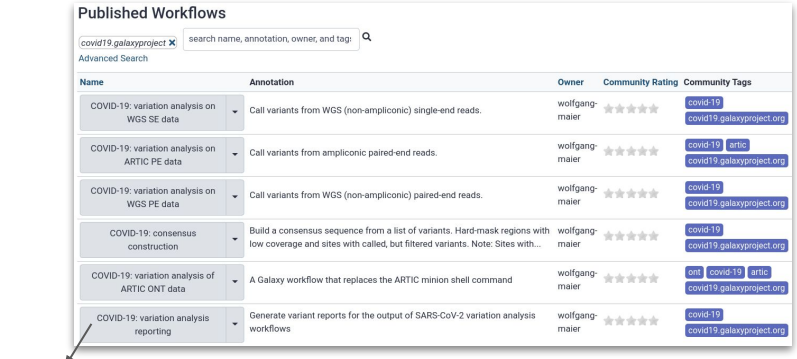
Galaxy knows how to direct user requests to cluster where the tools are running

More technical



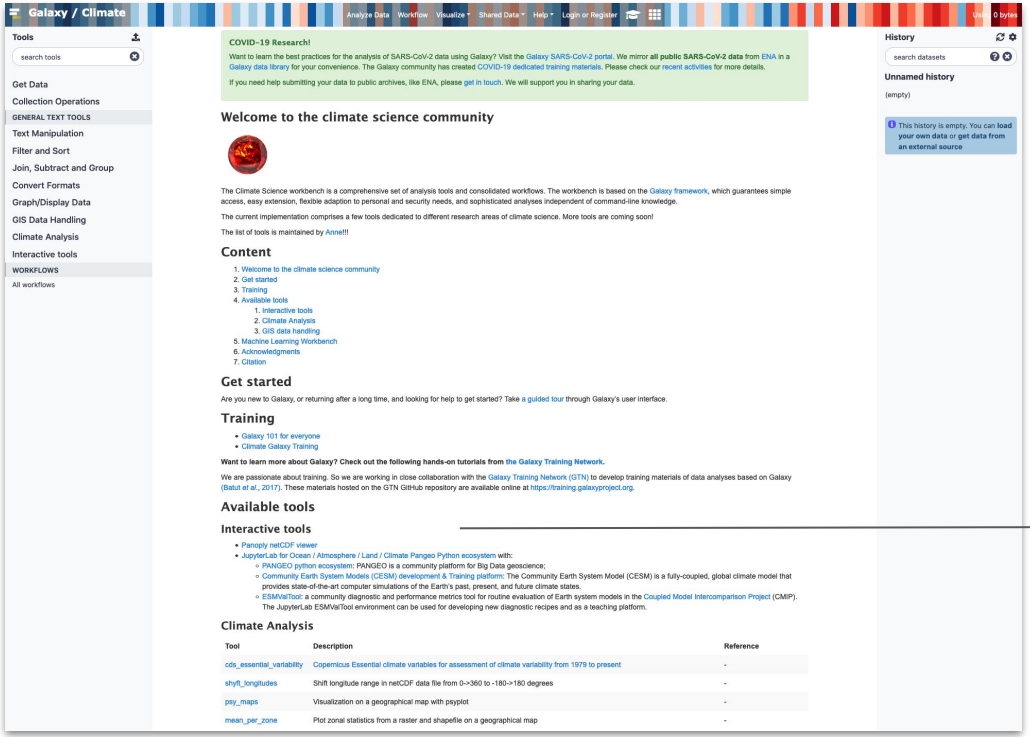
Interactive tools in COVID-19 variant calling analysis



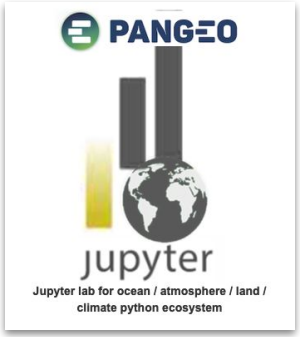


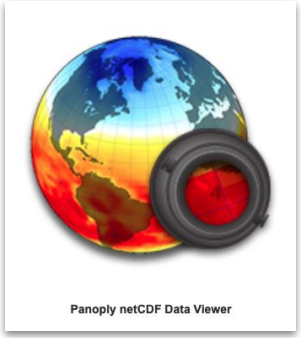
1 year ago → Now

Climate analysis on usegalaxy.eu



Interactive tools for climate





climate.usegalaxy.eu

JupyterLab for Ocean / Atmosphere / Land / Climate

PANGEO

jupyter

Jupyter lab for ocean / atmosphere / land / climate python ecosystem

jupyter + PANGEO

A community platform for Big Data geoscience

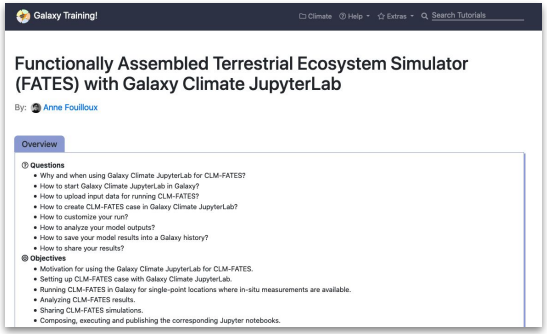
COMMUNITY EARTH SYSTEM MODEL CESM ®

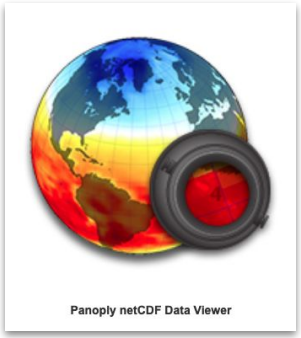
ESMValTool

Earth System Model Evaluation Tool

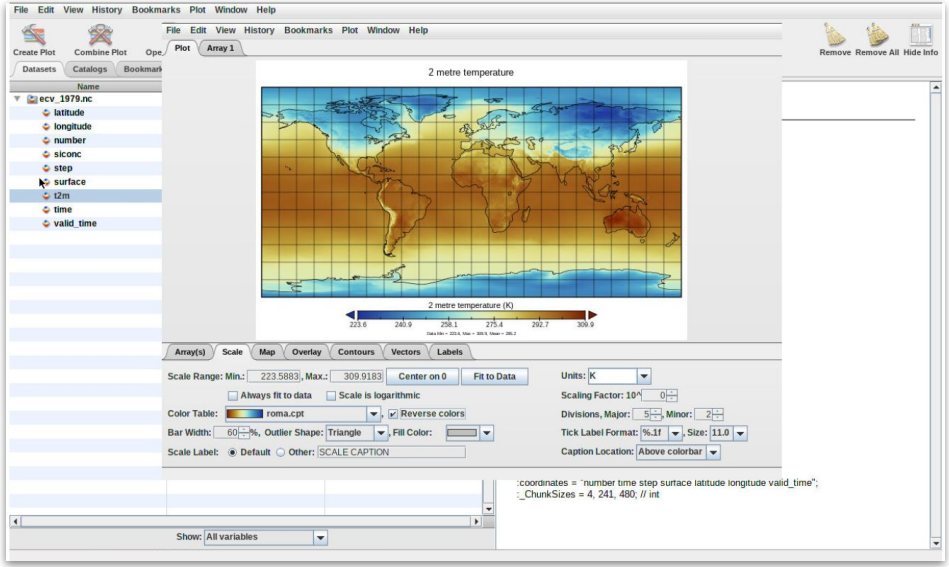
Use cases

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



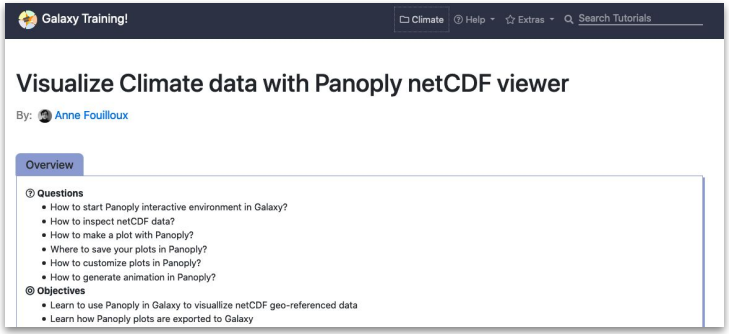


Panoply netCDF viewer



Use cases

- Data visualization
- Image saving
- Time series video creation



Ecology with Interactive Tools: 3 use cases

Inside a complete analysis

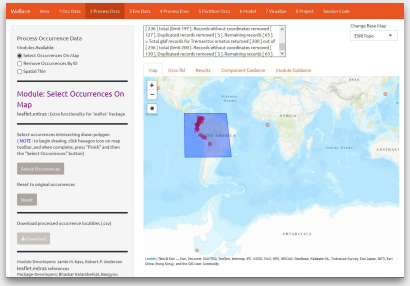
Biodiversity and environmental data collection

↓

Selection / Filtering

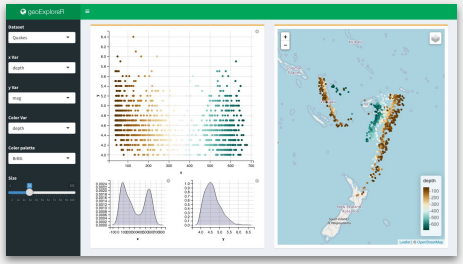
↓

Analysis via modelling
R Shiny Wallace



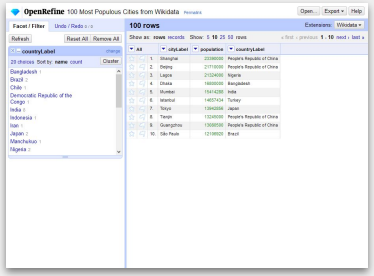
Data visualisation

Cartography + visualisation
R Shiny GeoExplorer



Data cleaning

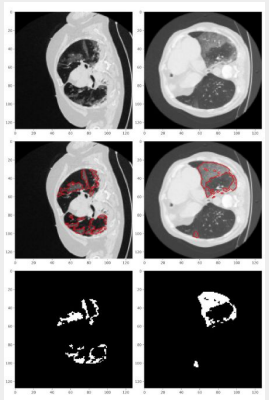
Data cleaning
OpenRefine



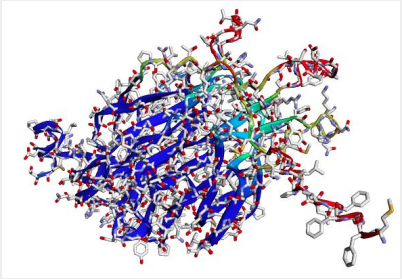
GPU-enabled VREs for everyone

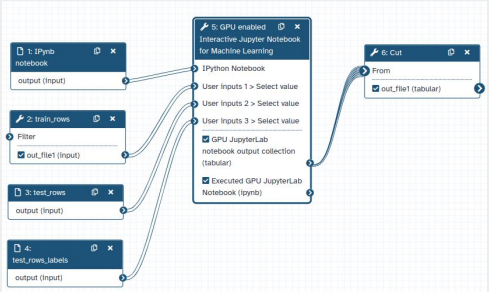
An accessible infrastructure for artificial intelligence using a Docker-based JupyterLab in Galaxy

Image segmentation of COVID-19 CT scans



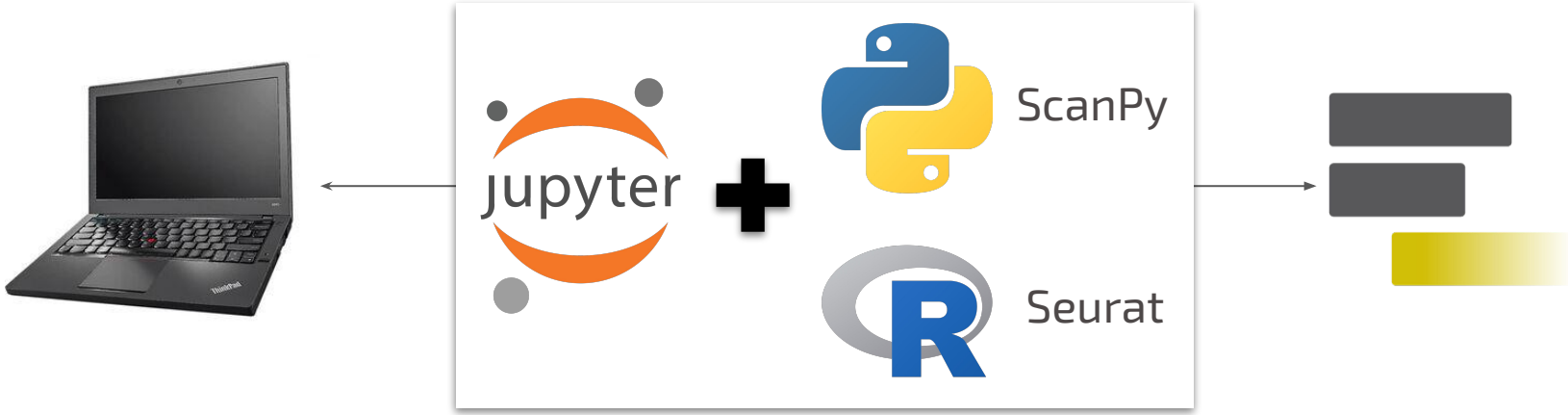
Prediction of 3D structures of proteins





Single-cell data analysis

When offline limitation (RAM, dataset size, etc)



Project Partners