1.1.1 Work package 3: Pulsar Network: Distributed heterogeneous compute

Explain the work carried out in WP3 during the reporting period giving details of the work carried out by each beneficiary/linked third party involved.

WP3 aimed to create and maintain a strong, production-ready distributed computing network across Europe, allowing scientific workflows to run in different compute centers, therefore leveraging state-of-art compute facilities and technologies across the EU.

To achieve this, WP3 relied on three core technologies: the Galaxy workflow manager, to provide access to data, tools, workflows, and compute resources; Pulsar, to fetch jobs to remote cluster and build the distributed computing network; and the CERN-VM FS, to share reference datasets and tools across all infrastructures.

To simplify the deployment of Pulsar endpoints and Galaxy instances, one of the main objectives of WP3 was the development of the Open Infrastructure: a set of Terraform and Ansible scripts, integrated with a Continuous Integration / Continuous Delivery system like Jenkins. This framework enables the automated deployment, configuration, and, most importantly, maintenance of new Pulsar endpoints and Galaxy instances, providing administrators with a comprehensive toolset to support their daily operations. The European Pulsar Network currently includes 13 endpoints and is used by six national Galaxy instances in addition to the European one (usegalaxy.eu).

Online meetings:

Since the beginning of the project, a monthly coordination meeting for Work Package 3 has been held on the first Tuesday of each month, for a total of 32 meetings. Each session followed a structured agenda, starting with a report from the WP leaders' meeting, followed by updates from all tasks. Particular attention was given to Task 3.3 (Pulsar network deployment) and Task 3.5 (usegalaxy.* operations), where system administrators regularly reported on the status of endpoint deployments and Galaxy instance operations. Meeting minutes were consistently recorded and made available to all participants.

Contributions:

ALU-FR, CNR, INFN, IISAS and EGI updated and tested the Pulsar Open Infrastructure for deploying and updating Pulsar endpoints. CNR and TUBITAK ULAKBIM contributed to the documentation. CNR developed the testing framework. CESNET developed and tested the TESP microservice. CNR, ALU-FR, CESNET, INFN, EGI, IISAS, TUBITAK ULAKBIM, Cyfronet, BSC and CNRS deployed and/or updated the Pulsar endpoints. ALU-FR and CNR developed and tested the Open Infrastructure framework for deploying and managing usegalaxy national instances. VIB contributed to the multi service deployment of Pulsar. ALU-FR, VIB, CNRS, CESNET, BSC-CNS updated and maintained their corresponding Galaxy instances. CNR deployed a prototype version of Italian usegalaxy instance.

- O3.1 Build a European wide job-scheduling network
- O3.2 Make Pulsar endpoints conform the GA4GH Task Execution Service standard
- O3.3 Deploying TRL-9 web services to access the Pulsar Network

Task 3.1 Develop and maintain an Open Infrastructure based deployment model for Pulsar endpoints (Aligned with objective O3.1)

This task focused on building and sustaining an Open Infrastructure (OI) that streamlines and scales the deployment of Pulsar endpoints over different settings. Building on the existing Pulsar Network deployment strategies, the aim was to make it simple for Consortium partners to start new nodes and handle growing compute needs while extending support to new scientific domains.

Already well-integrated with OpenStack and OpenNebula, widely used tools like Ansible, Packer, and Terraform helped automated deployment. The job also included expanding this assistance to container orchestration systems like Kubernetes and significant commercial cloud providers (AWS, Azure, Google Cloud). Integration with EOSC-compliant Authentication and Authorization Infrastructure (AAI) guaranteed compatibility with other EOSC services. This allows users to start Pulsar deployments straight from the Galaxy interface (in line with Task 4.1), therefore simplifying access to several cloud infrastructure.

September 2022- August 2023 summary of achievements:

Every Pulsar endpoint has several key components: an HTCondor-based compute cluster, CERN Virtual Machine File System (CVMFS) for tool and reference data access, Network File System (NFS) for internal data sharing, and the Pulsar application (which manages job execution and results exchange with Galaxy).

We improved the Open Infrastructure (OI) framework to enable automated Pulsar deployment on OpenStack-based clouds using Terraform and Ansible during the first reporting period. A separate GitHub repository offered documentation and ready-to-use deployment recipes for resource providers to configure Pulsar endpoints with a single command.

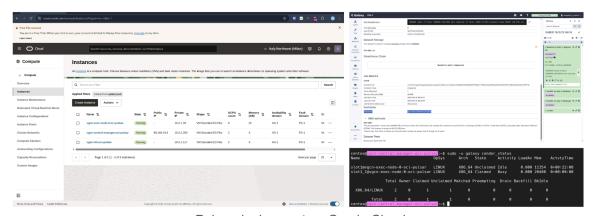
Except for HTCondor, which is now installed post-deployment owing to its changed authentication criteria, the deployment uses an updated version of the VGCN image now based on Rocky Linux 9, which bundles the required software. The HTCondor Ansible role was thus changed to support the most recent version (10.4.3). The configuration has been validated successfully across several partner systems (such as XXX), therefore verifying durability and repeatability.

September 2023- August 2025 Achievements:

Development efforts have been primarily directed toward two main goals: ensuring the long-term stability and maintainability of the European Pulsar infrastructure, and establishing a robust automated testing framework for continuous monitoring of endpoint availability. These activities are essential to support reliable operations and enable prompt intervention in case of failures. In this context, the following work has been carried out to improve and consolidate the Open Infrastructure and its associated components.

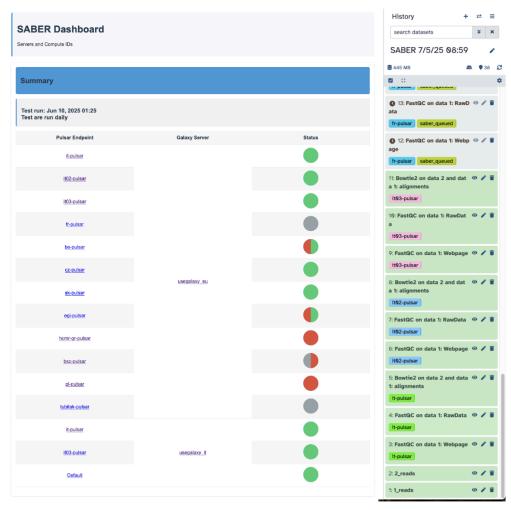
HTCondor has been upgraded to version 24.x, introducing minor but useful enhancements (<u>pull request</u>). Moreover, a key improvement is the adoption of a multi-Pulsar deployment model, which enables different Galaxy instances to share the same set of endpoints, improving resource efficiency and simplifying maintenance (<u>pull request</u>). The corresponding documentation is available <u>here</u>. At present, the Italian endpoints are jointly used by both UseGalaxy.eu and the national Galaxy server.

Deployment activities have also been successfully tested on commercial cloud environments. In particular, a fully functional deployment has been validated on Oracle Cloud Infrastructure (OCI), as illustrated in the accompanying figures. This result was achieved through the combined use of Terraform and Ansible, which ensure compatibility with a wide range of commercial cloud providers, highlighting the flexibility and portability of the infrastructure. The documentation is available here.



Pulsar deployment on Oracle Cloud.

Finally, to ensure service reliability and prompt failure detection, an automated testing infrastructure has been introduced. The system consists of a lightweight Python script (saber) orchestrated through a Jenkins pipeline that runs on a daily basis. Each test cycle executes two lightweight jobs—a FastQC and a Bowtie2 alignment on sacCer3 data—to keep runtimes reasonable given the number of endpoints involved. SABER is designed to systematically validate Galaxy endpoints using customizable workflows and job configurations. It supports per-instance settings and produces timestamped, traceable outputs, including HTML and Markdown reports. Configuration is managed via encrypted YAML files, offering Ansible Vault-like security. Future developments include integration with Grafana dashboards and support for parallel workflow executions. Continuous monitoring is available at https://monitor.usegalaxy.it.



SABER report and Galaxy history.

Task 3.2 Add GA4GH Task-Execution-Service (TES) API to Pulsar (Aligned with objectives O3.2, O2.2, O3.3)

We included the GA4GH Task Execution Service (TES) API to extend Pulsar's interoperability outside Galaxy. Although Pulsar now allows job submission from Galaxy via REST or message queues—with safe transmission of job definitions, input data, and return of results—the present API is closely linked to Galaxy.

Integrating the TES standard helped us to make Pulsar available to a larger spectrum of workflow management systems throughout the research ecosystem. By means of TES, this enables outside services to send tasks to Pulsar endpoints, therefore promoting more usage of the European Pulsar Network and complementing EOSC's objectives for open, reusable infrastructure.

September 2022- August 2023 summary of achievements:

We created an open-source task execution engine following the GA4GH TES specification (TES Standard) in line with our aim to assist TES-based job execution throughout the Pulsar network. Already compatible with workflow engines such Snakemake and Nextflow, the present implementation allows job submission to Pulsar nodes with REST-API access.

Using a test configuration including a standalone TESP server, backend database, FTP service, and a Pulsar instance, deployment has been confirmed on the Czech Galaxy Node. Using OAuth2 for authentication, the engine supports three data transfer protocols: S3, HTTPS, and FTP.

Though its development status limits some aspects, great effort was made to guarantee compatibility with the new Galaxy TES Runner. Although workflows can be started from Galaxy to Pulsar using TES, result transfer back to Galaxy is now inoperative. The Runner's lack of support for designating Docker images is a major restriction since it stops completely portable execution environments.

September 2023- August 2025 Achievements:

We continued to build on the foundations laid in earlier development of the TESP-API. The focus

has been on making the service more practical, stable and more TES compliant. A lot of this involved solving small-but-important problems and making sure things behave more reliably in real-world setups.

- We developed a test suite to validate the TESP-API's core functionality and to support more robust staging and un-staging of data. This has helped us catch edge cases early and streamline internal data handling.
- We implemented Docker Compose profiles, giving users the flexibility to run the full service stack or just the TESP-API and database containers. This makes development and deployment easier and more modular.
- The TESP-API now supports input and output parameters of type DIRECTORY broadening its compatibility with a wider range of workflows and use cases.
- With the help of José María Fernández, we identified some issues in how metadata is collected and exposed and a limitation in how the executor builds Docker run commands—specifically, that relying on sh -c could fail in minimal containers lacking a shell. We've since implemented a check that detects shell-dependent meta-characters and avoids using sh -c "" in simple cases, improving compatibility. These changes are in a testing phase and we are planning to integrate it with TESSAP in near future.
- We added support for BasicAuth authorisation to the TESP-API.

Task 3.3 Build a European-wide network of Pulsar sites (Aligned with objective O3.1)

This task focused on growing and sustaining a distributed network of Pulsar-enabled computing sites all across Europe. Participating partners provide local compute resources available via standardized Pulsar endpoints. Integrated with CVMFS, these endpoints guarantee efficient and repeatable access to tool containers and reference data—building on the infrastructure created by EOSC-Life.

Monitored via the public dashboard available at stats.galaxyproject.eu, every Pulsar site is connected to European Galaxy servers (as part of WP4). With active participation from key infrastructure partners such as EGI and EuroHPC, the main goal was to improve the readiness level of the Pulsar Network to TRL-9 by the end of the project.

September 2022- August 2023 summary of achievements:

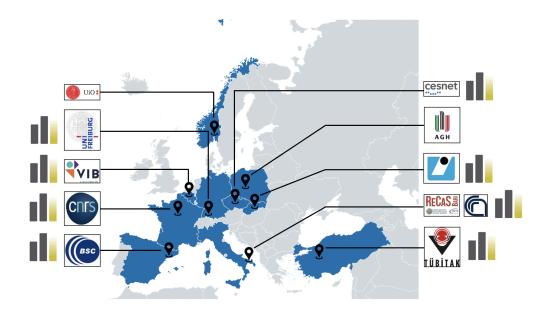
Several new Pulsar endpoints were deployed or updated across Europe building on the Open Infrastructure enhancements from Task 3.1:

- IT02 and IT03 in Italy [CNR]
- SK01 in Slovakia [IISAS]
- FR01 in France (GenOuest) [CNRS]
- CZ01 in Czech Republic [CESNET]
- EGI01 in Italy [EGI]

usegalaxy.eu's listed Pulsar sites are kept as infrastructure-as-code in a public repository.

September 2023- August 2025 Achievements:

Currently the European Pulsar Network encompasses 12 endpoints, created by the consortium partner, using the pulsar deployment repository. Moreover, one endpoint has been added by the HCMR institute for the FairEase (<u>Pull Request</u>) project.



Institution	Country	Endpoint name
ALU-FR	Germany	DE01
CNR	Italy	IT01, <u>IT02</u> , IT03 and IT04
CNRS	France	FR01
IISAS	Slovakia	<u>SK01</u>
CESNET	Czech Republic	<u>CZ01</u>
VIB	Belgium	BE01
EGI and INFN	Italy	EGI01
BSC	Spain	BSC01
TUBITAK ULAKBIM	Turkey	TUBITAK01
Cyfronet	Poland	<u>CFY01</u>
FairEase project	Greece	HCMR01

Task 3.4 Add TES support to WfExS (Workflow Execution Service) (Aligned with objectives O3.1, O2.2)

This task aimed to improve WfExS, a platform supporting the GA4GH TRS standard for fetching workflows from sources including WorkflowHub (WP2) that abstracts over several workflow management systems (WMS).

WfExS expanded as part of EuroScienceGateway to function as a pluggable compute backend, therefore enabling workflows to run across the European Pulsar Network. We worked on WfExS to utilize the GA4GH TES API (created in Task 3.2), guiding workflow engines like Nextflow and CWL to send tasks to distant Pulsar endpoints.

While also supporting the more general objectives of the project on workflow portability and reproducibility (Task 2.2), this task showed how other WMS can use the EuroScienceGateway infrastructure for task execution.

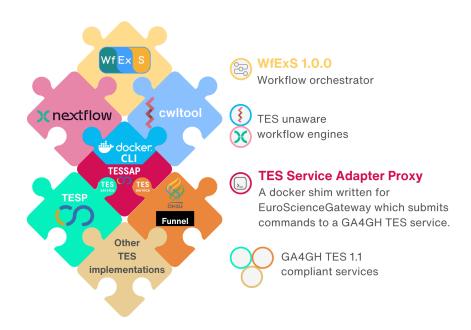
Nothing was mentioned in the 1st periodic report.

September 2023- August 2025 Achievements:

We have written TESSAP (https://github.com/inab/tessap), a docker to GA4GH TES service proxy, i.e. a docker client shim which submits commands to a GA4GH TES service.

WfExS-backend delegates workflow executions to the most appropriate supported workflow engine (currently either Nextflow or cwltool). Although Nextflow supports GA4GH TES in some degree, cwltool itself does not, and the same can happen to other supported workflow engines in the future.

But, all these workflow engines share a common feature: all of them support running their workflow steps using docker client (when the workflow has been written to do it so). The idea behind TESSAP is replacing the original docker client, maintaining both the parameter and behaviour compatibility as much as possible.



Not all the docker client commands have to be supported by TESSAP, because both nextflow and cwltool only need a subset of them in order to run commands within a container instance. These are the implemented or faked ones on version 0.7.1:

Implemented	Subcommand	Nextf low	cwltool	WfExS-backend
V	docker run	V	V	Ø
V	docker rm	V		Ø
V	docker stop	V		Ø
V	docker kill	V		Ø
V	docker ps			
V	docker pull		\	V
9>	docker stats		V	Ø
V	docker inspect			V
X	docker import		V	Ø
424	docker load		V	V
X	docker build		V	8

å. L	docker save		V
****	docker images		V
****	docker tag		V
324	docker rmi		V
\$25\$	docker version		V

So, most of previous subcommands from docker are already implemented (\checkmark) or faked (\diamondsuit). Other ones are going to be implemented or faked (\checkmark), and a few ones are not going to be even tried (\checkmark).

The implemented commands bypass the original implementation, they try to mimic the original commands implementation, through either forwarded calls to a set up GA4GH TES service, or API calls to the corresponding docker registries.

For other commands, the line is passed to the locally installed docker binary.

Task 3.5 Developing and maintaining national or domain-driven Galaxy servers (aligned with objectives O3.1, O2.1, O3.3)

Based on an Infrastructure-as-Code deployment model derived from the successful usegalaxy.eu approach, this task focused on the development and long-term maintenance of national or domain-specific Galaxy servers. Designed to support access to both local compute resources and the European Pulsar Network (Task 3.3), these instances have reached TRL-9 maturity (M3.3).

All participating instances worked together to assist a common catalogue of tools and workflows, which has been made discoverable via EOSC registries including bio.tools and WorkflowHub (Task 2.1), therefore guaranteeing a consistent user experience across the federation. A maturity model is being defined to guide deployment and sustainability, including factors such as:

- Pulsar Network integration
- Sharing of container images and reference data over CVMFS
- Federated AAI systems, such as LS Login, provide user authentication.

Partners participated in a common user support system depending on a forum-based approach to meet community needs, resolve technical problems, and harmonize feedback throughout the network.

September 2022- August 2023 summary of achievements:

Using a standarized Infrastructure-as-Code approach, national Galaxy instances were deployed and maintained during the first reporting period. This strategy not only allowed the simple instantiation of new services but also offered a sustainable, maintainable framework for updating current deployments. Below is a list of participating partners' contributions:

- Leading the development of the Open Infrastructure, usegalaxy.eu (ALU-FR) upgraded Galaxy from versions 22.5 to 23.1 and moved to the Total Perspective Vortex (TPV) scheduler. Several Pulsar endpoints were combined, and a new user interface component was added to let users choose distant resources. To assist testing operations across work packages, a test replica of the EU production server was also built.
- Using CINECA resources, IT (CNR) started running the usegalaxy.it instance. Established a
 strong test environment using TPV configuration, CVMFS for reference data, HTCondor
 scheduling, and PostgreSQL replication. The PON CNR.BiOmics and ELIXIRxNextGenIT
 initiatives have provided hardware resources for the production instance. We have deployed
 and tested a Pulsar endpoint.
- 3. Upgraded the usegalaxy.be test instance considerably, including OS migration (CentOS to Rocky Linux), Galaxy upgrade (21.01 to 23.0), and Postgres upgrade (v9 to v15). Configured TPV, RabbitMQ, and Pulsar integration as well. Infrastructure automation is in progress and will be published.
- 4. Maintained and improved the usegalaxy.fr instance, now operating Galaxy 23.0 with TlaaS and tools supporting the WP5 Biodiversity use case. The server is linked to a strong Slurm cluster and runs a big user base. TPV migration is under progress.
- 5. Ongoing use of the usegalaxy.no infrastructure with both production and test stacks. The test environment has been enhanced and internal knowledge of Ansible-based deployment

- maintenance was increased with new personnel.
- 6. Connected to e-INFRA CZ's PBSPro system with dedicated GPU-enabled compute resources, CZ (CESNET) deployed a national Galaxy instance (usegalaxy.cz) with TPV and Pulsar integration. A dual AAI system serves national users as well as ELIXIR.
- 7. OpenStack resources were used to run usegalaxy.es. Featuring Slurm scheduling, CVMFS reference data, and a big toolset, the instance is fully functional. Plans are in place to deploy a Pulsar endpoint and expand the configuration using more Slurm worker nodes.

September 2023- August 2025 Achievements:

EU [ALU-FR]: usegalaxy.eu continues to lead in delivering robust, scalable, and sustainable Galaxy infrastructure across Europe. The usegalaxy.eu instance has been successfully upgraded to the latest Galaxy 25.0 release, ensuring access to the most up-to-date features, security improvements, and user enhancements.

To support scalability and meet growing user demands, the team has integrated additional Pulsar endpoints into the European Pulsar Network. In parallel, multiple subdomains were developed and launched to serve specific research communities, facilitating tailored tool environments and simplified onboarding.

Several new functionalities have been enabled, including:

- Bring Your Own Storage, Compute, and Data capabilities, enhancing user flexibility and autonomy
- Integration of domain-specific tools and interactive tool environments to support advanced, real-time analysis workflows
- Integration of RO-Crates, RSpace, eLabFTW, and Dataverse, improving data management, documentation, and reproducibility
- Ongoing support for federated AAI systems, now extended to include EGI Check-in, ELIXIR LS Login, DataPLANT, and NFDI

The infrastructure has undergone a comprehensive upgrade, including significant enhancements to the compute, storage, and software stack. Additionally, we have acquired next-generation HPC, GPU, and storage hardware to support the growing Galaxy Europe user base, which now exceeds 130,000 active researchers.

To further streamline deployment and support the growth of federation, new community-maintained OpenStack images have been released to simplify Pulsar endpoint provisioning. In parallel, Grafana dashboards have been developed to improve monitoring and observability across the entire infrastructure.

To support outreach and community engagement, the team has conducted numerous training events and actively represented the EuroScienceGateway project at leading conferences, including EGI 2024 (https://galaxyproject.org/news/2024-10-10-egi2024/) and EGI 2025, where we hosted exhibition booths to showcase developments and foster collaboration.

IT [CNR]: CNR has deployed the Italian Galaxy instance, UseGalaxy.it (https://usegalaxy.it), leveraging Open Infrastructure. Currently, a development instance has been implemented using cloud resources provided in-kind by CINECA (https://www.cineca.it), an ELIXIR-ITALY partner, while the production instance is operational at ReCaS-Bari. UseGalaxy.it infrastructure is described in the blog post: https://galaxyproject.org/news/2024-11-23-usegalaxy-it-starting/.

- https://usegalaxy.it domain activated.
- Update nfs share from v3 to v4 to fix performance issues, preventing job to run correctly.
- Galaxy updated to 24.2 version: https://github.com/usegalaxy-it/infrastructure-playbook/pull/11
- Life Science AAI login enabled: github.com/usegalaxy-it/infrastructure-playbook/pull/9/commits/7f0e51985536cfa9446b0d7e9755 5347a3c648c9
- Celery/flower configured: https://github.com/usegalaxy-it/infrastructure-playbook/pull/8
- Groups and quota configuration done.
- Tools installation and/or updates: https://github.com/usegalaxy-it/usegalaxy-it-tools
- Moving Galaxy services to a physical server. This is due to the storage configuration, which is on a different network compared to the cloud.
- The infrastructure automation framework and documentation are hosted on Github:
 - Usegalaxy-it github repository: https://github.com/usegalaxy-it
 - Usegalaxy-it operations documentation: https://usegalaxy-it.github.io/documentation
- New hardware resources to be used by ELIXIR-IT have been acquired in the context of the ELIXIRxNextGenIT RRF project (https://elixir-italy.org/project/elixirxnextgenit). Some of those

resources will be dedicated to service the incoming UseGalaxy.it server, whose main production instance will be deployed at the ReCaS-Bari data center.

BE [VIB]: VIB has continued to deploy and upgrade the Belgian Galaxy instance, <u>UseGalaxy.be</u> (https://usegalaxy.be), on the OpenStack cloud provided by the Flemish Super Computing Center (https://www.vscentrum.be/). During the project <u>usegalaxy.be</u> was upgraded in the following ways:

- Galaxy updated to version 24.2
- Bring your own data feature enabled (and hierarchical to distributed galaxy object store config)
- Increased availability through deploying multiple gunicorn services, allowing a rolling restart
- Set up telegraf/influxdb/grafana for increased monitoring
- Increased tool offering through automated tool installation of the iuc tools (https://github.com/galaxyproject/tools-iuc)
- Use iRODS as the default Galaxy object store
- Improved backup and recovery strategy

FR [CNRS]: CNRS has continued maintaining and developing the French Galaxy instance, <u>UseGalaxy.fr</u>. This instance is hosted on the national French Bioinformatics Institute computing infrastructure, giving access to significant HPC resources (4700 cores, 64Tb of RAM, 11 GPUs).

In addition to already deployed features (e.g. TlaaS, Life Science Login authentication), the main updates achieved during the period on this instance include:

- Successive upgrades up to version 24.2
- CVMFS stratum 1 mirror (internal use)
- Celery/RabbitMQ/Flower services for internal Galaxy tasks and Pulsar endpoint connection
- Total Perspective Vortex for job scheduling
- Galaxy Labs (domain specific homepages, synchronised with other instances)
- Activation of recent Galaxy features (RO-Crate export, Bring Your Own Data, ...)

The French instance is following the Galaxy Open Infrastructure best practices, all the administration tasks being performed using public GitLab repositories (https://gitlab.com/ifb-elixirfr/usegalaxy-fr/).

<u>UseGalaxy.fr</u> is now delivering a mature service to the community, and counts 10400 user accounts and a total of 6.2M jobs (3000 active users and 2.5M jobs in the last year). The management of this instance is performed by a national workgroup, "Galaxy France", that also promotes the use of Galaxy in the French ecosystem, by organising regular meetings, connecting Galaxy users, developers and admins across France.

NO [UiO]: >> PLEASE ADD <<

CZ [CESNET]: CESNET has advanced UseGalaxy.cz to a mature production instance of Galaxy that is kept up to date and uses distributed deployment architecture taking advantage of hardware virtualization.

It runs jobs using exclusively Pulsar-enabled destinations and Singularity images leveraging robust PBS job scheduler service of CESNET's metacentrum.cz which includes the national ELIXIR CZ compute cluster and a dedicated GPU machine for AlphaFold 2 jobs among other resources. On the Galaxy's side it uses Total Perspective Vortex for scheduling optimizations. Available tools more or less mirror the usegalaxy.eu toolset and are automatically updated every week. Job running is rigorously tested using automated job executors which are integrated to a wider monitoring using the checkmk testing tools.

For local indices and other shared data it accesses Galaxy project's global CVMFS. In terms of authentication it is connected to multiple AAI providers including LS-Login and e-infra.cz. Notably we have configured TIAAS for training and also support export/import of RO-Crates including in combination with repository systems Zenodo and InvenioRDM.

All of the infrastructure is maintained following Galaxy standards as a public repository with Ansible playbooks at https://github.com/CESNET/usegalaxy. To promote our work we've presented at multiple regional conferences and meetings including ELIXIR CZ annual conference, Czech PhD bioinformatics conference, brno.bio meetup, and others.

ES [BSC]: BSC has deployed the Spanish Galaxy instance, <u>usegalaxy.es</u>, currently available as a development instance. The server is hosted and maintained by the Barcelona Supercomputing Center (BSC), with the domain <u>usegalaxy.es</u> already activated and functional. Galaxy has been updated to version 24.0. The installed toolset is aligned with those on other major Galaxy instances, like <u>usegalaxy.eu</u>. One of the main developments carried on during this period has been a complete platform cloud migration, moving the entire Spanish Galaxy ecosystem to a newer version of OpenStack (ncloud to ncloud2).

The GitLab repository hosting the infrastructure codebase and related automation available at: https://gitlab.bsc.es/inb/usegalaxy/galaxy.

Deliverables and Milestones

- D3.1 Operations documentation on the Open Infrastructure deployment
- D3.2 Publication on the Pulsar Network, integrated in workflow management systems
- **M3.1** Pulsar network is TRL-9: operational
- M3.2 Demonstrated job submission via the WfExS to the Pulsar Network
- M3.3 National Galaxy servers reaching TRL-9 (operational in environment)