Horizon Europe (HORIZON)
HORIZON-INFRA-2021-EOSC-01

Periodic Report

Technical Report (Part A)
Technical Report (Part B)
Financial Report

Version 1.1
01 May 2023

IMPORTANT NOTICE

What is the Periodic Report?

The Periodic Report/Final Report is the pre-condition for receiving payments; it must be submitted through the EU Funding & Tenders Portal Grant Management System by the Coordinator within 60 days after the end of the reporting period.

The Report is divided into a technical and financial report.

The Technical Report consists of 2 parts:

- Part A contains structured tables with project information
- Part B is a narrative description of the work carried out during the reporting period.

Part A is generated by the IT system. It is based on the information which you enter into the Portal Continuous and Periodic Reporting modules.

Part B needs to be uploaded as PDF on the Technical Report (Part B) screen. The template to use is available there.

The Financial Report normally consists of:

- the individual financial statements (Annex 4 to the GA) for each Beneficiary
- a summary financial statement
-
- a certificate on the financial statements (CFS) (if threshold reached).

The Financial Report is generated by the IT system on the basis of the financial information entered into the Periodic Reporting module (and any other documents uploaded, e.g. CFS).

How to prepare and submit it?

The Periodic Report must be prepared by the Consortium in the Continuous and Periodic Reporting modules and then be submitted by the Coordinator.

The Continuous Reporting module is always open and can be updated at any moment during the project (submit deliverables, report on milestones, etc.). It automatically feeds Part A of the Periodic Report.

The Periodic Reporting module is opened after the end of the reporting period. It allows you to:

- download and upload the Part B of the Technical Report (upload only by the Coordinator)
- complete their financial statements on-line (each Beneficiary for themselves and their Affiliated Entities)
- consolidate the individual financial statements into a summary financial statement (Coordinator)
- submit the Periodic Report (Coordinator).

Make sure that all the information in the Continuous Reporting module is updated *before* 'locking the periodic report for review'. Updates entered after this step will be included in the Periodic Report of the following period (if any).

This document is tagged. Be careful not to delete the tags; they are needed for the processing.

DEFINITIONS	
Background	Any data, know-how or information whatever its form or nature, tangible or intangible, including any rights such as intellectual property rights, that is (i) held by beneficiaries prior to their accession to the action; and (ii) identified by the beneficiaries in a written agreement as needed for implementing the action or for exploiting its results.
Communication	Communication on projects is a strategically planned process that starts at the outset of the action and continues throughout its entire lifetime, aimed at promoting the action and its results. It requires strategic and targeted measures for communicating about (i) the action and (ii) its results to a multitude of audiences, including the media and the public and possibly engaging in a two-way exchange.
Critical risk	A critical risk is a plausible event or issue that could have a high adverse impact on the ability of the project to achieve its objectives. Level of likelihood to occur: Low/medium/high The likelihood is the estimated probability that the risk will materialise even after taking account of the mitigating measures put in place.
Clustering activities	The project has participated in joint activities with other research projects (funded either by the EU or not, such as the D&E Booster, etc) to promote its research results as part of a portfolio of projects.
Deliverable	A report that is sent to the granting authority, providing information to ensure effective monitoring of the project. There are different types of deliverables (e.g. a report on specific activities or results, data management plans, ethics or security requirements).
Dissemination activities	The public disclosure of the results by any appropriate means (other than resulting from protecting or exploiting the results), including by scientific publications in any medium.
Exploitation	The utilisation of results in further research activities other than those covered by the action concerned, or in developing, creating and marketing a product or process, or in creating and providing a service, or in standardisation activities, or feeding back into policy making activities.
Impacts	Wider effects on society (including the environment), the economy or science, enabled by the outcomes of R&I investments in the long term. *Example:* “Airports increase their maximum passenger capacity by 15 % and the passenger average throughput by 10 %, leading to a 28% reduction in infrastructure expansion costs. This is achieved thanks to the use of an advanced forecasting system developed in the project.”
Key results (KER)	A Key result (KER) is a main result which has been prioritised due to its high potential for exploitation and further scientific, societal or policy use.
Milestone	Control points in the project that help to chart progress. Milestones may correspond to the achievement of a key result, allowing the next phase of the work to begin. They may also be needed at intermediary points so that, if problems have arisen, corrective measures can be taken. A milestone may be a critical decision point in the project where, for example, the consortium must decide which of several technologies to adopt for further development.
Objectives	The goals of the work performed within the project, in terms of its research and innovation content. This will be translated into the project’s results. These may range from tackling specific research questions, demonstrating the feasibility of an innovation, sharing knowledge among stakeholders on specific issues. The nature of the objectives will depend on the type of action, and the scope of the topic.

Outcomes	The expected effects, over the medium term, of projects supported under a given topic. The results of a project should contribute to these outcomes, fostered in particular by the dissemination and exploitation measures. This may include the uptake, diffusion, deployment, and/or use of the project's results by direct target groups.
Owner	The person/entity which has generated the results.
Pathway to impact	Logical steps towards the achievement of the expected impacts of the project over time, in particular beyond the duration of the project. A pathway begins with the projects' results, to their dissemination, exploitation and communication, contributing to the expected outcomes in the work programme topic/ JU work programme, and ultimately to the wider scientific, economic and societal impacts of the work programme destination/JU work programme.
Results	What is generated during the project implementation. This may include, for example, know-how, innovative solutions, algorithms, proof of feasibility, new business models, policy recommendations, guidelines, prototypes, demonstrators, databases, trained researchers, new infrastructures, networks, etc.
SDGs	Sustainable Development Goals
Technology Readiness Level	See HE Work programme General Annexes section B

TECHNICAL REPORT (PART B)

COVER PAGE

Part B of the Technical Report must be downloaded from the Portal Technical Report (Part B)/Termination Report screen, completed and then assembled and re-uploaded as PDF on that screen.

Note:

For EIC Accelerator actions: Please use this reporting template only for the final report. For additional prefinancing requests, this template should be replaced by the PPT used for the progress meeting (i.e. upload the PPT in place of the additional prefinancing report).

PROJECT	
Project number:	[101057388]
Project name:	[EuroScienceGateway]
Project acronym:	[EuroScienceGateway]

REPORTING PERIOD ⚠ *Please note that you must report on the entire reporting period.*	
RP number:	*Final*
Duration:	from [01/09/2022] to [31/08/2025]

#@PER-REP-HE@#
#@PRO-GRE-PG@# [This document is tagged. Do not delete the tags; they are needed for the processing.]

1. EXPLANATION OF THE WORK CARRIED OUT AND OVERVIEW OF THE PROGRESS

Include an overview of the project results towards the objective of the action in line with the structure of the Annex 1 to the Grant Agreement including summary of deliverables and milestones. In the technical description below, please avoid repeating information that is already present in part A of the report (in continuous reporting). For projects under topics indicating the need for the integration of social sciences and humanities, explain the role of these disciplines in the project so far.

(No page limit per work-package but the report should be concise and readable. Any duplication should be avoided).

This part presents a brief overview of the work done and the general progress achieved over the EuroScienceGateway (ESG) project (M1-M36). It emphasizes the key accomplishments toward achieving the project's specific objectives, as stated in the Description of Action (DoA), and it is organized accordingly.

From Life Sciences and Climate Research to Astrophysics and Material Sciences, data-driven research is still transforming the way scientific discovery is carried out across fields. These fields depend more and more on cooperative infrastructures and complex data analysis pipelines. At the same time, there is a rising need for the FAIR (Findable, Accessible, Interoperable, Reusable) sharing of not only data, but all digital research outputs. This approach promotes reproducibility, innovation, and successful reuse of scientific findings by supporting Open Science.

To simplify access to European computing infrastructures, Workflow Management Systems (WMS) are becoming a cornerstone technology. These systems enable a consistent, open, and smooth method to large-scale data processing. Without requiring deep technical expertise, Galaxy, the workflow platform at the core of EuroScienceGateway, enables researchers to compose, share, and run complex workflows. Supporting scalable, user-friendly, and FAIR-compliant data analysis, it acts as an abstraction layer over high-performance and cloud-based resources.

EuroScienceGateway aimed at establishing an efficient, interoperable, and configurable workflow-based gateway for European researchers. Throughout the course of the project, the consortium has made significant progress in establishing a production-level infrastructure integrating national and thematic service providers into the larger European Open Science Cloud (EOSC) ecosystem. Coordinated technical development, community involvement, service integration, and policy alignment initiatives have helped to address project goals as this report highlights.

1.1 Objectives

Please list the specific objectives for the project as described in section 1.1 of the DoA; Please provide a short summary of progress towards the achievement of each of the project objectives. Highlight significant activities in support of these achievements. Please provide clear and measurable details; report on objectives not fully achieved or not on schedule.

A vision of an open and collaborative digital environment that empowers researchers throughout Europe was the main objective of the EuroScienceGateway project. Four main project objectives have been used for achieving this vision, each of which is connected to specific, quantifiable results. These goals ensure coherence and alignment in both technical development and strategic implementation by acting as a connecting framework for the work done across all Work Packages.

Objective 1: Accessible e-Infrastructure resources for European scientists to enable pioneering data-driven research across scientific domains.

In areas like Life Sciences, biodiversity, Climate Science, Astrophysics, Material Sciences, and other data-intensive sectors, EuroScienceGateway aimed to close the gap

between the complexity of e-Infrastructure use and the pragmatic demands of researchers. Although national and institutional institutions and research entities enjoy more availability of high-performance computing (HPC) and cloud resources, technical constraints and limited access often keep them separate from domain-specific research projects.

EuroScienceGateway project has established a workflow-based gateway integrated with the Pulsar job distribution system and FAIR-compliant workflow tools (WorkflowHub, RO-Crate, metadata standards) centered around the Galaxy platform to address this issue. These elements taken together are being implemented as a scalable, flexible, production-ready service (TRL-9) that meets various scientific requirements all over Europe (and the globe).

Key technical milestones were achieved during the first reporting period: While integration frameworks like "Bring-Your-Own-Storage" (BYOS) and "Bring-Your-Own-Compute" (BYOC) were completed, Pulsar endpoints were effectively deployed and upgraded across partner sites. Early adopter communities in Biodiversity, Climate Science, Materials Science, and Astrophysics started onboarding onto the Galaxy platform, therefore offering useful verification of the gateway approach (e.g., Deliverable 5.2).

Extensive training programs reaching more than 4,000 researchers as well as new Galaxy administrators and developers helped to promote adoption. By guaranteeing reusability and adoption by other communities, documentation and open-source code outputs (e.g., Deliverable 3.1 (Nicotri et al. 2024)) help to eliminate the need for scattered, domain-specific solutions.

More technical milestones are achieved during the second reporting period, too. More partner Pulsar sites deployed and came online (e.g. TUBITAK01). Extensive documentation for the installation and maintenance for Pulsar endpoints have been authored, and the infrastructure and process for Pulsar endpoint deployment is greatly improved to both improve the deployment process and make maintaining active Pulsar sites easier.

Objective 2: Support the varieties of analysis types and diverse usage patterns through efficient and smart job distribution to appropriate and sustainable infrastructures.

By enabling the efficient execution of different jobs across a variety of computing environments, EuroScienceGateway aimed to assist a wide spectrum of scientific analyses. Scientific analyses differ greatly in their infrastructure needs from batch to interactive workflows and from compute-intensive to data-heavy tasks. Jobs have to be dynamically matched to appropriate resources—factoring in not only hardware requirements but also energy consumption, financial costs, and data locality if they are to guarantee best performance, sustainability, and cost-efficiency (see Deliverable 4.2 (Azab, De Geest, et al. 2025; Azab, Srikakulam, et al. 2025)).

The project has created middleware that abstracts access to partner-managed infrastructures together with smart job scheduling techniques to fulfill this objective (Milestone 4.1, Milestone 4.2, Deliverable 4.1 (Pedersen et al. 2024) and Deliverable 4.2 (Azab, Srikakulam, et al. 2025)). The implementation of the "Bring-Your-Own-Compute" (BYOC) and "Bring-Your-Own-Storage" (BYOS) capabilities (M4.1) during the first reporting period resulted in a major technical milestone. These let local resources be flexibly integrated into the larger platform ecosystem, enabling specialized or institution-specific computing arrangements.

At the same time, first concrete steps are taken toward including data locality into the job scheduling system—a vital element for performance optimization and minimizing needless data transfers, particularly for data-intensive workloads. This feature is a major improvement over traditional scheduling systems, which often ignore how data movement affects efficiency and cost.

Objective 3: The application of FAIR principles to workflows and adoption of FAIR Digital Objects to stimulate reusable and reproducible research and enable the EOSC Interoperability Framework.

EuroScienceGateway contributed to open and reproducible research by including FAIR concepts into the entire lifecycle of computational workflows. Reproducibility has become easier to achieve across research fields as workflow systems, notebooks, and containerized tools have been increasingly used. We aimed to make sure that processes are recorded as FAIR Digital Objects (FDOs)—and are interoperable, well-documented, and reusable across infrastructures and disciplines.

We set the groundwork for including FAIR practices into data analyses during the first reporting period. This included recording of Galaxy workflow executions, formalizing them as FAIR Computational Workflows using RO-Crate, and publishing them in the WorkflowHub registry. With more than 40 workflows made publicly available[1]—offering a rich and varied collection of reusable workflows for the scientific community, Galaxy became a major contributor.

Our input went beyond publishing workflows. We started to influence best practices for the larger EOSC ecosystem by including FDOs and related metadata standards into Galaxy and collaborating with technologies like RO-Crate and FAIR Signposting (Clark-Casey and Soiland-Reyes 2022). These changes directly support the EOSC Interoperability Framework[2] and have already affected related initiatives in ELIXIR and the FDO Forum—showing EuroScienceGateway's contribution in promoting FAIRness in computational research[3].

Objective 4: Adoption of the EuroScienceGateway by researchers in diverse scientific disciplines.

EuroScienceGateway seeked to promote widespread use within scientific communities and its success is quantifiable by statistics on how researchers from various fields have accessed and reused data, tools, and workflows (Objective 5.1, Objective 5.2, Objective 5.3, Task 1.6, Task 5.1, Task 5.2, Task 5.3, and Task 5.4). The results and usage of the project demonstrates how our gateway simplifies scientific discovery in many areas by means of proactive engagement with scientific use cases and communities (WP5) and dissemination (WP1).

During the first reporting period, we effectively launched and carried out the Dissemination and Exploitation Plan (D1.2[4]), boosting visibility via a committed website[5], engaged social media channels[6], and conference attendance. Targeted training sessions (Galaxy Training Academy[7,8,9]) to assist onboarding and use complemented community-building plans. During the ESG project, new communities have formed on Galaxy Training Network and existing communities curated further training materials[10].

The EuroScienceGateway, through its cornerstone platform Galaxy, has been used increasingly throughout the project. Bioinformatics is still a significant user base, but adoption has spread into new areas including proteomics, structural biology, biodiversity,

---

[1] https://workflowhub.eu/projects/166#workflows
[2] https://eosc.eu/roadmap/bringing-ro-crate-to-trl-9/
[3] https://training.galaxyproject.org/training-material/topics/fair/
[4] https://doi.org/10.5281/zenodo.10055554
[5] https://galaxyproject.org/projects/esg/
[6] https://mstdn.science/@galaxyproject, https://bsky.app/profile/galaxyproject.bsky.social, https://www.linkedin.com/company/galaxy-project/
[7] https://gallantries.github.io/video-library/events/smorgasbord3/
[8] https://training.galaxyproject.org/training-material/events/galaxy-academy-2024.html
[9] https://training.galaxyproject.org/training-material/events/2025-05-12-galaxy-academy-2025.html
[10] https://training.galaxyproject.org/

climate science, materials science, and astronomy.
At the time of writing this report, the European Galaxy server had more than 132,000 registered users spread across 30 community-maintained subdomains[11]. At least 19 other EOSC or other European initiatives have also used EuroScienceGateway, underlining the platform's increasing relevance and influence all across Europe. These successes show that EuroScienceGateway is well on track to promote a lively, domain agnostic user base—one that will keep growing by continuous community involvement and operational service delivery.

#@WRK-PLA-WP@#

1.2 Explanation of the work carried out per WP

Include a table or description of how the project has achieved each objective of the period.

The following section describes the work done under each Work Package (WP) throughout the project. Every WP overview starts with a short description of its objectives, main tasks, meetings conducted, and progress achieved including the particular contributions of partner organizations. A thorough report on accomplishments by tasks follows the summary.

1.2.1 Work Package 1: Project Management, Coordination, Dissemination

Explain the work carried out in WP1 during the reporting period giving details of the work carried out by each beneficiary/affiliated entity involved.

O1.1 Establishment of functions for project management.
O1.2 Organisation and execution of project management, managing board and project annual meetings.
O1.3 Development and implementation of a risk management plan to ensure project objectives are achieved.
O1.4 Develop and periodically update a data management plan.
O1.5 Establishment of a long-term sustainability model.
O1.6 Disseminate the project outcomes within well-established networks and scientific communities.
O1.7 Provision of new content and delivery of training for the deployment of the EuroScienceGateway Open Infrastructures

Summary

WP1 covered all aspects of project administration and project management to ensure effective reporting, and achievement of the project aims. Albert-Ludwigs-Universität Freiburg (ALU-FR) took on project coordination and has been responsible for the technical, financial and administrative management on a day-to-day basis and its interactions with the overall management structure. For the European Commission (EC), ALU-FR has been the official contact point. Decision-making processes, as well as dissemination, implementation and exploitation issues, have been documented in a consortium agreement that has been updated during the project if necessary.

Task 1.1 Establishment and implementation of the Project Management Unit (M1-36) (Aligned with objective O1.1)
Task lead: ALU-FR

A Project Management Unit (PMU) is established to oversee the EuroScienceGateway project's administration, management, and coordination, ensuring timely delivery of documents, effective communication among partners, and the organization of meetings and training events. The PMU also ensures the project stays within its budget, follows EC requirements for cost control, and monitors any changes or risks to the project.

ALU-FR led the establishment of the Project Management Unit (PMU) for the EuroScienceGateway project, overseeing key tasks such as financial and administrative management, internal communication, stakeholder engagement, monitoring, and

11 https://stats.galaxyproject.eu/d/bejvb2fpdi96ob/eu-factsheet?orgId=1&from=now-6h&to=now&timezone=browser

reporting to the European Commission (EC). The PMU defined key project roles and governing bodies, including the General Assembly for decision-making and the Executive Board for project direction and oversight. A Management Team supported day-to-day operations, organizing meetings and managing project tasks. The PMU created a project manual (ESG information hub) to establish procedures for efficient execution and submitted the Data Management Plan (DMP) (see Deliverable 1.1 (Erxleben, Schaaf, D'Anna, et al. 2023)) and Dissemination and Exploitation Plan (D&E Plan) (see Deliverable 1.2 (Erxleben, Schaaf, Marchis, et al. 2023)) as the first two deliverables. Effective communication channels were set up, including chat rooms[12], mailing lists[13], and online meetings, ensuring smooth coordination and timely delivery of milestones. The PMU also managed the project's budget, ensuring cost control and compliance with EC requirements.

The project has consistently adhered to the project management unit's initial plan throughout its execution, ensuring that all activities align with the established timelines and objectives. After the submission of the first periodic report, a follow-up meeting was conducted with the EU office. During this meeting, the feedback provided was overall positive, and several valuable recommendations were presented to further enhance project outcomes. These suggestions were carefully reviewed and promptly acted upon to ensure the continued success of the project. The details will be provided later in this report.

Some of the partners needed to transfer funds from one budget category to another to ensure that the project activities proceeded as planned. This change did not necessitate a formal amendment to the Grant Agreement, and the project officer was duly informed of the adjustments, ensuring transparent communication.

Additionally, the monthly meetings continued as per the established schedule, fostering regular updates and collaboration among partners. All deliverables were submitted on time, and the milestones set for the project were achieved without any complications or delays. Partners were also continuously encouraged to upload their achievements to the project portal, ensuring that progress was well-documented and accessible.

Furthermore, the partners actively shared their key developments and success stories through blog posts[14], contributing to the visibility of the project's impact and forming a strong sense of engagement within the wider community. This structured approach has allowed the project to remain on track and fully aligned with the expectations outlined at the outset.

Budget Reallocation and Reporting Update

A budget reallocation, representing 2% of the total project budget, was approved by the Project Officer. This internal adjustment allowed for the continuation of key project activities without impacting the overall work plan or deliverables. As the reallocation did not alter the scope or objectives of the project, no formal amendment to the Grant Agreement was required.

In the course of reviewing the first periodic report, some issues were identified that require additional justifications from a few partners. The consortium worked closely with those partners to provide the necessary clarifications and ensure full compliance with reporting requirements.

Continuous Reporting on the EC Portal

To support timely and accurate updates to the European Commission's Continuous Reporting portal, all partners were provided with an Excel template to regularly document key developments from their respective work packages, including publications, results, and other relevant outputs.

---

[12] eurosciencegateway:matrix.org

[13] esg-admin@lists.galaxyproject.org, esg-consortium@lists.galaxyproject.org, esg-wp2@lists.galaxyproject.org, esg-wp3@lists.galaxyproject.org, esg-wp4@lists.galaxyproject.org, esg-wp5@lists.galaxyproject.org

[14] https://forum.eosc.eu/, https://galaxyproject.org/projects/esg/news/ and https://galaxyproject.org/projects/esg/events/

The coordinator contributed to this process by extracting relevant information from the project's blog posts and adding it to the shared table. The coordination team conducted monthly reviews to ensure that the table remained up to date and that all reportable items were reflected in the portal in a timely manner.

Meeting with the EU & follow up on recommendations:

Recommendation 1: Further attention is needed to the approach to security in the light of the EU's changing position and the increasing need to protect EU IP against hostile actors.

How we addressed it: We are constantly hardening our infrastructure, responding to CVEs and streamlining our process in the ISO 27001 framework[15].

Recommendation 2: Additional efforts should be made to promote the website and introduce the work to a wider set of stakeholders.

How we responded: The EuroScienceGateway project and the Galaxy platform have been actively promoted across a variety of research communities and events. These outreach efforts aimed to raise awareness, increase adoption, and build synergies with complementary initiatives. Notable events and communities where the project was presented include: EOSC Symposium 2024, EOSC Winter School 2024, EOSC Winter School 2025, EGI2024, EGI2025, Galaxy Community Conference (GCC2023), GCC2024, Galaxy and Bioconductor Community Conference 2025 (GBCC2025), JOBIM 2024, JOBIM 2025, Galaxy Imaging Hackathon 2025, European Galaxy Days (EGD) 2023, EGD 2024, EGD 2025, PASC23 Conference, Swiss SKA days 2023, Galaxy Admin Training 2023, Swiss SKA days 2023 and many more. Besides, we regularly posted updates on the EOSC Forum and actively participated in the monthly EOSC communication meetings, where project-relevant events were shared with over 20 other project managers. As a result of this engagement, several other EOSC projects (such as EOSC-Nordic[16], EOSC-Pillar[17], EOSC-Life[18], AquaInfra[19], Blue-Cloud[20], Data Commons[21], FAIR-EASE[22], OSCARS[23], BY-COVID[24], EOSC4Cancer[25], FAIR2ADAPT[26], Skills4EOSC[27], EGI-ACE, AgroServ, ELIXIR Converge[28], European Genomic Data

---

15 https://galaxyproject.org/news/24-01-2025-27001-certification/
16 https://galaxyproject.org/news/2022-10-17-eosc-pillar-agreement/
17 https://galaxyproject.org/media/scientific-collab-agreement-eosc-pillar-eosc-life.pdf
18 https://training.galaxyproject.org/training-material/hall-of-fame/eosc-life/
19 https://galaxyproject.org/news/2023-11-14-esg-w-p5/ and https://galaxyproject.org/news/2025-04-07-aquainfra-eosc-and-galaxy/
20 https://blue-cloud.org/events/advancing-open-science-interoperable-data-access-and-cloud-based-workflows
21 https://galaxyproject.org/news/2025-04-07-data-commons/
22 https://galaxyproject.org/news/2023-05-21-fair-ease-euro-science-gateway/, https://galaxyproject.org/events/2023-05-training-fair-ease/, https://galaxyproject.org/news/2025-04-02-fair-ease-brest-hackathon-2025-news-post/ , and https://galaxyproject.org/community/sig/earth/
23 https://galaxyproject.org/news/2024-02-06-oscars-project-starting/, https://galaxyproject.org/news/2024-10-21-eoscxnfdi-joint-symposium/, and https://galaxyproject.org/community/sig/image-analysis/
24 https://galaxyproject.org/news/2024-09-16-by-covid-eol/ and https://training.galaxyproject.org/training-material/events/tracks/gta2024-bycovid.html
25 https://galaxyproject.org/news/2022-09-30-eosc-kickoff-meeting/
26 https://galaxyproject.org/news/2025-05-20-jupytergis/ and https://training.galaxyproject.org/training-material/events/tracks/gta2025-climate.html
27 https://training.galaxyproject.org/training-material/hall-of-fame/skills4eosc/
28 https://galaxyproject.org/news/2022-10-19-rdm-galaxy/, https://galaxyproject.org/news/2023-06-24-elixir-annual-report-2022/, and https://training.galaxyproject.org/training-material/hall-of-fame/elixir-converge/

Infrastructure (GDI), Erasmus Gallantries[29], BioGenEU, BioNT[30], HealthyCloud[31]) have adopted Galaxy as part of their infrastructure or workflows.

To further increase the project's visibility, promotional materials such as stickers, stands, posters, factsheets and brochures content were distributed at multiple events. Additionally, consortium members contributed to community papers and publications (see complimentary continuous reporting tables) that highlighted the project's goals, achievements, and potential for collaboration.

EuroScienceGateway is now featured in several strategic frameworks: it is listed on the EOSC Macro Roadmap[32], recognized as one of the services under the National Research Data Infrastructure (NFDI)[33], and serves as the EOSC Life Science Node[34]. Furthermore, user engagement has significantly exceeded expectations, with over 31,000 more users than initially anticipated and reaching over 131,000 registered users[35].

Recommendation 3: Consider requesting a training needs analysis and associated training plan for each of the use cases. Consider a user requirements survey.

How we addressed it: Training needs were assessed through multiple channels, including the GTA/Smörgåsbord survey and targeted questions in registration forms and online discussions. In addition to that, we contributed to EOSC Opportunity Area 5 (OA5), focusing on skills and training needs at a strategic level. We also engaged with the Skills4EOSC project, though coordination challenges limited its impact compared to more mature projects like GTN. While there is currently no unified ESG learning pathway, contributing communities such as climate science and ecology have developed their own tailored training paths. New training demands continue to emerge as new communities onboard.

Recommendation 4: A recommendation for the next period will be to develop further the integration of existing data repositories beyond EOSC, such as the EHDS, into the Galaxy, to refer to interoperability guidelines and standards developed by other EOSC projects such as FAIR-IMPACT (WP8) and to promote further the project to other research communities and also at the level of ESFRI (ICT working group, EOSC taskforce) to other Research infrastructures, beyond those within EOSC. Further efforts are also needed to promote the services within communities already integrated in the project, such as Astrophysics- for which the ESFRI networks could play a positive role.

How we addressed it: While direct integration with the European Health Data Space (EHDS) falls outside the scope of this grant, the project has actively engaged with other initiatives such as EOSC4Cancer, where Galaxy serves as a core component, reinforcing our alignment with broader health data interoperability goals.

Over the past two years, we have significantly expanded Galaxy's integration capabilities, enabling by default access to 38 external data repositories—ranging from general-purpose platforms like Zenodo to domain-specific ones such as BERD from the Social Sciences and Humanities field. We have also introduced a "Bring Your Own Data" (BYOD) concept, allowing users to integrate and work with data from their own data using services such as OwnCloud, NextCloud, Amazon S3 Storage, and any other 3rd-party repositories, further enhancing flexibility and interoperability (Deliverable 4.1**(Pedersen et al. 2024)**).

---

[29] https://galaxyproject.org/events/2023-06-12-train-the-trainers/ and https://training.galaxyproject.org/training-material/news/2021/10/12/data-science.html

[30] https://galaxyproject.org/news/2023-09-28-biont-workshop/, https://galaxyproject.org/news/2023-11-21-2nd-biont-workshop/, and https://training.galaxyproject.org/training-material/hall-of-fame/biont/

[31] https://galaxyproject.org/news/2021-03-09-healthy-cloud/

[32] https://eosc.eu/eosc-about/eosc-macro-roadmap/?_search=eurosciencegateway

[33] https://galaxyproject.org/news/2023-01-03-nfd-i4-bioimage/ and https://galaxyproject.org/news/2020-02-07-data-plant/

[34] https://www.eosc-life.eu/news/fair-ease-and-eurosciencegateway-start-3-year-collaboration/

[35] https://stats.galaxyproject.eu/d/bejvb2fpdi96ob/eu-factsheet?orgId=1&from=now-6h&to=now&timezone=browser&viewPanel=panel-2

Galaxy continues to adopt and promote EOSC-wide interoperability guidelines and standards, including those emerging from projects like FAIR-IMPACT. These efforts support data reuse across domains and contribute to Galaxy's recognition as a recommended interoperability service by ELIXIR[36]. Promotion of our services to new and existing research communities, including ESFRI networks, remains an ongoing priority—particularly in domains such as astrophysics, where further engagement is already underway.

Task 1.2 Organization and execution of all project management and partner meetings (M1-36) (Aligned with objective O1.2)
Task lead: ALU-FR

The Project Coordinator and the Project Management Unit organized, led, and documented all project management meetings, prioritizing virtual meetings and scheduling them to coincide with annual general meetings to minimize costs and foster productive discussions among Work Package Leads and research infrastructure personnel.

The Project Coordinator, together with the Project Management Unit (PMU), organized, led, and documented all Project Management Unit meetings, Work Package meetings, and meetings of the Executive Board and General Assembly. Work Package Leaders monitored their respective tasks through regular meetings and reported progress in Managing Board meetings, highlighting completed and ongoing work, input needed from other Work Packages, deliverables and milestones, and any delays or risks, with suggested mitigations. To minimize costs, most meetings were held virtually, and project management body meetings were scheduled to coincide with the annual General Assembly and the European Galaxy Days[37], fostering productive discussions and networking. In addition to ESG partner meetings, the team participated in Galaxy community calls and working group meetings. The PMU also supported the organization of various training events, including a significant Galaxy admin workshop in Gent, Belgium[38], which covered key topics for setting up high-performance Galaxy instances.

Project Governance and Strategic Meetings

In addition to the regular monthly Work Package Leaders' meetings and work package–specific discussions, the project has maintained a strong governance structure through the organization of annual General Assemblies. These took place in October 2023 in Freiburg and in October 2024 in Berlin, each strategically aligned with major community events to foster collaboration, knowledge exchange, and the dissemination of best practices.

The 2023 General Assembly was held alongside the European Galaxy Days in Freiburg[39], while the 2024 meeting coincided with the EOSC Symposium in Berlin[40]. During these meetings, General Assembly members reviewed project progress, evaluated key outcomes, and contributed to the development of strategic directions and decisions for the consortium.

To further support effective coordination and oversight, the Executive Board convened online every three months. These meetings played a vital role in ensuring the smooth progression of the project and addressing any emerging risks or challenges in a timely manner.

Strategic Planning and Extended Collaboration

[36] https://galaxyproject.org/news/2023-12-14-elixir-rir-for-galaxy-europe/
[37] https://galaxyproject.org/events/2025-10-01-egd2025/, https://galaxyproject.org/events/2023-10-egd/, and https://galaxyproject.org/events/2022-10-egd/
[38] https://galaxyproject.org/events/2023-admin-training/ and https://training.galaxyproject.org/training-material/topics/admin/
[39] https://galaxyproject.org/events/2023-10-egd/esg/
[40] https://galaxyproject.org/events/2024-04-25-esg-annual-meeting

During the General Assembly meetings, members engaged in strategic discussions focused on enhancing the visibility and impact of the project. This included the development and refinement of marketing and communication strategies, the identification of key events and conferences for representation, and the exploration of future opportunities for engagement and growth.
A significant emphasis was placed on strengthening the project's relationship with the broader EOSC ecosystem. Members actively discussed approaches to deepen collaboration within the EOSC network and explored mechanisms to foster sustainable partnerships beyond the project's duration.

In terms of outreach and collaboration, the project has exceeded initial expectations (see Continuous reporting Part A). Engagement extended beyond the originally anticipated scope, successfully connecting with a wider range of stakeholders, projects, and communities than initially proposed. This achievement underscores the project's strong positioning within the European research infrastructure landscape and its commitment to fostering an open, interoperable, and collaborative environment.

list

- Open collaboration with the US Galaxy community working on materials science at Oak Ridge National Lab. (https://www.ornl.gov/staff-profile/gregory-watson)
- Open collaboration with the Extreme Photonics Application Centre at the Science and Technology Facilities Council in the UK. (https://www.clf.stfc.ac.uk/Pages/EPAC.aspx)
- Open collaboration with the x-ray absorption spectroscopy (XAS) community. OSCARS project combining Galaxy and the Cross Domain Interoperability Framework to process XAS data. (https://oscars-project.eu/projects/cdif-4-xas-describing-x-ray-spectroscopy-data-cross-domain-use)
- Supported EOSC AquaInfra for Galaxy tool development; Deployment of tools and workflows on the European Galaxy instance; Support for Galaxy related Training Material
- Collaboration with Earth Biogenome Project and affiliated projects (Vertebrate Genome Project, European Reference Genome Atlas, Biodiversity Genomics Europe, ATLASea) for the co-construction of genome annotation resources (tools, workflows, trainings) and support in their adoption by the community.
- Support FAIR-EASE for tool development; Deployment of tools and workflows on the European Galaxy instance; joined workshops, presentations and demos; Support for Galaxy related Training Material
- Codeveloped training materials with Skills4EOSC
- Offering support to OSCARS projects, like OSCARS-Fidelis
- Cooperation with EOSC Data Commons in executable RO-Crates and Galaxy. RO-Crate is the chosen specification to implement the concept of Package for Processing Datasets in EOSC Data Commons. Galaxy is chosen as one of the target execution environments to execute the Package for Processing Datasets via the EOSC Data Player.
- WorkflowsRI
- EOSC United (officially starts in October 2025) is demonstrating the ESG results, like BYOC and BYOS via the EOSC EU Node.

Final Event Planning

The final project event is scheduled to take place in Freiburg after the official project end date (October 2025), maintaining continuity with previous General Assemblies and aligning with the European Galaxy Days[41]. In accordance with funding regulations, all associated costs will be incurred and accounted for before the project's official end date.

Task 1.3 Quality and risk management (M1-M30) (Aligned with objectives O1.3 and O1.4)
Task Lead: VIB
Task Members: all partners

Scope:

[41] https://galaxyproject.org/events/2025-10-01-egd2025/

The task will create and continuously update a Data Management Plan (DMP)[42] that incorporates ethical and legal compliance, with a focus on open-source tools, Open Access publications, and registration of generated artefacts in EOSC-supported registries. Additionally, a management contingency plan will be developed to assess and mitigate risks throughout the project.

September 2022- August 2023 summary of achievements:

This task developed the Data Management Plan (DMP) for the project, addressing ethical and legal compliance while incorporating components from each Work Package. The DMP, delivered in February 2023, outlined the use of open-source tools, Open Access publications, and the registration of generated artefacts in EOSC-supported registries, with regular updates throughout the project. Potential risks were identified, and the task ensured the quality of work by hosting technical work in public repositories on Github and contributing to upstream repositories. Feedback from training and conference participants was routinely collected to improve performance[43]. Although most risks from the grant proposal did not occur, two medium-likelihood risks—delay in recruiting qualified personnel and loss of key personnel—were managed by temporarily assigning roles to experienced staff. Despite these challenges, all deliverables and milestones were successfully submitted on time. The document was updated throughout the project when the situation required it.

Task 1.4: Development of a sustainability model and community assessment (M1-M30) (Aligned with objective O1.5)
Task Lead: EGI
Task Members: all partners

Scope:

The task aims to establish a long-term sustainability model for the EuroScienceGateway by conducting market research to identify additional target communities, estimate user volume, and evaluate stakeholder involvement. The Open Infrastructure model and global community alignment will enhance the network's resilience, ensuring capacity recovery and increased reliability with more partners worldwide.

September 2022- August 2023 summary of achievements:

During this reporting period, meetings and preliminary exchanges with stakeholders and other RIs were held to discuss the development of the sustainability analysis for the Galaxy platform.
September 2023- August 2025 Achievements:

The Sustainability model has been created by the EGI Foundation and is online available here: add link. The Executive Summary has been included in this Report.

Galaxy is a free, open-source system for analyzing data, authoring workflows, training and education, publishing tools, managing infrastructure, and more. This report has examined Galaxy's sustainability and strategic value through the lenses of desirability, feasibility, and viability. The findings make a compelling case that Galaxy represents a high-impact, low-risk investment opportunity for funders, policy makers, and research institutions seeking to support open science and accelerate discovery.

Galaxy Fulfills a Critical Need (Desirability): Researchers around the world have flocked to Galaxy as data analyses become more complex and compute-intensive. Usage metrics for Galaxy demonstrate extraordinary demand. For instance, the European Galaxy server grew from just a few dozen monthly active users in 2014 to over

[42] https://doi.org/10.5281/zenodo.10054551
[43] https://galaxyproject.org/news/2023-03-21-microbial-survey/, https://galaxyproject.org/news/2025-04-17-gat-survey/, https://galaxyproject.org/news/2023-09-28-biont-workshop/#applications-and-pre-workshop-survey, https://galaxyproject.org/news/2023-11-21-2nd-biont-workshop/, https://training.galaxyproject.org/training-material/news/2024/12/06/spoc_cofest.html, https://galaxyproject.org/news/2023-10-13-egd-2023-the-wrapup/, and https://galaxyproject.org/news/2023-03-29-physalia-assembly-annotation/

6,000 monthly users by 2024, with scientists executing on the order of a million analysis jobs per month on this server alone. This explosive growth reflects Galaxy's unique ability to make advanced science accessible – users with little to no coding skills can perform reproducible analyses through Galaxy's user-friendly web interface. Importantly, Galaxy's users don't just sign up and vanish; a significant proportion return regularly, indicating strong user satisfaction and "stickiness." As of 2024, nearly half of all active users each month are repeat users who come back on multiple days. Such loyalty underscores that Galaxy has become an indispensable tool in the researcher's toolbox. The clear conclusion is that there is massive and sustained demand for Galaxy's capabilities. Funding Galaxy is therefore funding a platform that thousands of scientists are already avidly using (and many more would, given the opportunity), meaning investments directly map to supporting real-world research productivity.

Proven Team and Infrastructure (Feasibility): Galaxy isn't just popular; it's robust and well-managed. Over two decades, what began as an academic project has matured into a global collaboration with a strong engineering backbone. The Galaxy developer community is large, distributed, and experienced, which assures that the platform can be maintained and improved over the long haul. Dozens of contributors drive Galaxy's development, preventing over-reliance on any single individual and keeping bus factor risks low. The project has established governance structures and an inclusive culture that welcomes new contributors, ensuring a steady inflow of talent and ideas. From a technical standpoint, Galaxy's infrastructure has proven it can scale reliably. Even as usage grew by orders of magnitude, the system's reliability increased – recent years show low error rates (~93% of jobs succeed) despite the enormous workload, thanks to continuous performance tuning and resource expansion. Major public Galaxy servers operate on high-performance computing resources, and the team has experience managing these at scale. For policy makers, this means Galaxy has the feasibility aspect well covered: Money invested in Galaxy is handled by a competent community and has been translated into dependable services for scientists. There is no "single point of failure" in personnel or technology. Galaxy, in effect, has the governance and infrastructure akin to a well-run international organization or a mid-sized tech company – but one that operates for the public good.

High Impact and Broad Adoption (Viability): Galaxy's value is evident in the breadth of its adoption and the depth of its impact on science. The platform has been cited in over 13,000 research publications to date, a clear indicator that it is enabling discoveries across genomics, medicine, agriculture, environmental science, and more. This is not an isolated community project; it is a globally recognized resource. Galaxy is integrated into national research infrastructures on several continents. In Europe, it is a key part of the ELIXIR ecosystem, and it has received funding from Horizon Europe to bolster its services. In Australia, Galaxy is part of the national life science computing infrastructure. In the United States, it's funded by the NIH as an open resource for all scientists. Such institutional adoption speaks volumes: organizations charged with advancing science have chosen to fund Galaxy. The ability to deliver societal benefit in times of need (like during COVID pandemic) shows Galaxy's agility and importance beyond routine research. In terms of viability, the bottom line is that Galaxy has become embedded in the fabric of modern science. It delivers continuous value, and its widespread network of users and contributors ensures that it will remain relevant. Supporting Galaxy yields a multiplier effect: it empowers a whole community of researchers, educators, and even citizen scientists, amplifying the impact of money invested.

A Low-Risk, High-Reward Investment: From an investment perspective (be it a grant, institutional support, or other funding), Galaxy stands out as a smart, strategic choice. The risk is low because the project is long-established and has proven sustainability – it's not going to disappear overnight, and it has robust mechanisms to manage personnel turnover and technology evolution. The core team has navigated growth for years, and the software's open-source nature means even in the unlikely event that official support waned, the community could continue the project (a safety net not available to proprietary ventures). On the other hand, the potential rewards are very high: by funding Galaxy, supporters enable thousands of downstream research projects and innovations. Few investments in science infrastructure have such a broad enabling effect. It is akin to funding a road or a power grid for data-driven analysis – it may not itself produce a Nobel Prize, but it may be the very platform on which Nobel-winning research is built. Galaxy lowers barriers to advanced analysis, leveling the playing field so that even researchers without big budgets or specialized skills can perform cutting-edge computational

research. This democratization of science accelerates discovery and innovation across the board. For funding agencies with mandates to foster open science, reproducibility, and collaboration, Galaxy is a flagship that already embodies these principles and will carry them forward with additional support.

Task 1.5 Outreach and dissemination with tailored communication and branding (M1-36) (Aligned with objective O1.6)
Task Lead: ALU-FR
Task Members: all partners

Scope:

The success of the EuroScienceGateway project relies also on effectively communicating its
goals, objectives, and achievements, especially to end-users, by leveraging the established networks of consortium partners. A dissemination and exploitation plan is to be
developed and executed, including strategies for updating the project website, maintaining a
blog, participating in conferences, using social media channels, and fostering community building through success stories to reach 100,000 users in the next three years.

September 2022- August 2023 summary of achievements:

WP1 focused on establishing and delivering a project-wide communication strategy for the EuroScienceGateway (ESG) project to increase visibility, inclusivity, and coherence of ESG's messaging. It managed both internal and external communications with an emphasis on end-users, leveraging established networks within the consortium and targeting new channels. The project's communication efforts included developing and maintaining a project website[44], creating blog posts[45], and utilizing social media platforms to reach broader audiences. ESG branding and materials, such as a logo[46] and communication templates were created to ensure consistent visibility. Furthermore, WP1 engaged in outreach activities, including citizen science projects and conferences (see technical report part A), to disseminate project results and foster engagement with diverse stakeholders.

September 2023- August 2025 Achievements:

EuroScienceGateway has played a pivotal role in expanding and strengthening the Galaxy network across Europe and beyond. Prior to the project's inception, our collaborations were primarily concentrated within a limited set of stakeholders, including MoU with EOSC-Nordic and EOSC-Pillar[47]. The Galaxy Freiburg team has also been collaborating in the EOSC-Life project.

However, over the course of the three-year project, EuroScienceGateway has substantially broadened its reach. We are now actively collaborating with a wide array of institutions, research infrastructures, and strategic initiatives, including [see list below], marking a significant leap in both scale and impact. This growth reflects not only the project's success in fostering meaningful scientific partnerships, but also its commitment to advancing open, interoperable, and sustainable research ecosystems.

https://eosc.eu/horizon-europe-projects/

- AquaInfra is using Galaxy
- Blue-Cloud is offering Galaxy
- DataCommons Galaxy is part
- FAIR_ease is using Galaxy
- We are part of OSCARS + there are OSCAR projects that are relying on Galaxy
- BY-COVID have used Galaxy, and we have been part of it

[44] https://galaxyproject.org/projects/esg/
[45] https://galaxyproject.org/projects/esg/news/ and https://galaxyproject.org/projects/esg/events/
[46] https://github.com/usegalaxy-eu/branding/tree/master/euro-science-gateway
[47] https://galaxyproject.org/news/2022-10-17-eosc-pillar-agreement/

- EOSC4Cancer is using Galaxy
- FAIR2ADAPT is using Galaxy
- Skills4EOSC is complicated, but I think they contributed to the GTN

Participation in Conferences and Events

The number of conferences and events attended throughout the duration of the project has significantly exceeded initial projections. While this resulted in additional travel-related expenses (to be detailed in the financial section of the report), it also brought substantial benefits in terms of visibility, engagement, and strategic development opportunities for EuroScienceGateway.

Through presentations, posters, and hands-on workshops, the project was prominently represented at a wide range of high-impact scientific and community events. These engagements have not only amplified the project's reach but also fostered meaningful connections with new stakeholders and potential collaborators.

Below is a list of key events where EuroScienceGateway was actively represented:

· EOSC Symposium/Madrid (booth)

· Winter School 29th of January 2024 until the 2nd of February 2024 (travel time included) in Thessaloniki, Greece

· https://www.ds4s.ch/ Bern, 11-12 April 2024

- NFDI CoRDI/Karlsruhe (talks & posters)
    - 720 (!) onsite participants
    - Two talks (GTN/TIaas, Galaxy and RDM) given, 3-4 ESG posters (ITs, Conda/containers/bots) presented. Blog posts will be written.
    - Several RO-Crate mentions/posters. See also https://www.researchobject.org/ro-crate/outreach.html
    - Leyla Jael Castro, Stian Soiland-Reyes, Dietrich Rebholz-Schuhmann (2023): **RO-Crates Meets FAIR Digital Objects**. [poster]
    - Sebastian Schaaf, Anika Erxleben-Eggenhofer, Bjoern Gruening (2023): **Galaxy and RDM**. [poster/slides??]
    - Carole Goble, …, Frederik Coppens (2023): **EOSC-Life Workflow Collaboratory for the Life Sciences**. [slides]

Task 1.6 Capacity building via developer and admin training (M7-30) (Aligned with objectives O1.7 and O5.3)
Task Lead: ALU-FR
Task Member: VIB

Scope:

Training is a key part of the dissemination strategy to make European computational infrastructure accessible, with the Galaxy Training Network (GTN) supporting the development and sharing of reusable materials. The project will extend the GTN resources by mentoring new communities, organizing Codefest events for tool developers, and collaborating with EOSC projects to provide training using EuroScienceGateway as a Training Infrastructure (TIaaS).

September 2022- August 2023 summary of achievements:

Training has been a key part of the dissemination strategy to make the European computational infrastructure more accessible to researchers, with the Galaxy Training Network (GTN) providing reusable materials and mentoring new communities. During the reporting period, the GTN expanded its resources with more tutorials, reaching over 1.7 Milion visitors, and provided training through events like Codefests, workshops, and online courses, including the large-scale Smörgasbord event. The project also organized specialized training for tool developers and collaborated with EOSC projects to offer EuroScienceGateway as a Training Infrastructure (TIaaS), while supporting over 16,000

trainees through more than 400 events. Additionally, the team provided hands-on workshops, including a specialized admin training session in 2023, and facilitated global collaboration with Galaxy administrators and developers to promote sustainability beyond Europe.

September 2023- August 2025 Achievements:

- Galaxy Training Academy global online training event 7-11 October. with over 2800 registrations was successful (blogpost) https://training.galaxyproject.org/training-material/events/2025-05-12-galaxy-academy-2025.html#overview

Deliverables
D1.1 EuroScienceGateway data management and contingency management plan **(M6, Public) D1.2** Dissemination and exploitation plan **(M6, Public)**
D1.3 Final summary of EuroScienceGateway main achievements, impacts, key results, sustainability data management and exploitation plans **(M36, Public)**

1.2.2 Work Package 2: Stimulate FAIR and Reusable Research

Explain the work carried out in WP2 during the reporting period giving details of the work carried out by each beneficiary/linked third party involved.

Summary

WP2 aims to embed strong FAIR principles into EuroScienceGateway by integrating workflows and their related data being published, shared, and reused as FAIR Digital Objects (FDOs). Encapsulated as RO-Crate, these FDOs will contain linked resources, logs, and metadata, so qualifying them for findability, reusability, and long-term archiving. Published via EOSC services and registries like OpenAIRE, these items will be promoted by means of focused outreach to both current and new user communities, therefore improving visibility and adoption. Establishing WorkflowHub—a RO-Crate-compliant FAIR workflow registry—as the go-to registry for workflows across disciplines within EOSC. A key goal is to raise this from its present TRL-7 to a production-ready TRL-9 service. WP2 works closely with the INFRA-01-EOSC-01-05 CSA project as well to guarantee compatibility with persistent identifiers (PIDs) and metadata systems, so enhancing the interoperability of research results across scientific disciplines.

Online meetings:
WP2 held monthly meetings throughout the length of the project on the last Thursday of the month, with occasional exceptions during holidays, totalling 28 meetings. Each session followed a structured agenda with minutes recorded in a shared document, including reports from each subtask and the WP leaders meeting. There were some periods with additional meetings for task forces for specific subtasks, such as the WorkflowHub knowledge graph milestones M2.3 and M2.4.

Contributions:
UNIMAN developed WorkflowHub toward TRL-9 operational standards. UNIMAN and EGI contributed to integration of WorkflowHub with EOSC infrastructure. UNIMAN, EGI, and UiO contributed to collaborations on standards and FAIR Digital Objects with other international initiatives, including EOSC projects. UNIMAN, BSC, UiO, VIB, and ALU-FR contributed to the RO-Crate specification, the Workflow RO-Crate and Workflow Run Crate profiles, and their integration into WorkflowHub, Galaxy and WfExS. ALU-FR, CNRS, EPFL, UKRI, UNIMAN, and VIB contributed workflows to the curated EuroScienceGateway collection. UNIMAN, EPFL, and BSC developed the WorkflowHub knowledge graph. UNIMAN hosted the WorkflowHub Publishers and Journals forum.

O2.1 Bringing FAIR workflows into EOSC through the EuroScienceGateway
O2.2 Support reusable and reproducible workflows

O2.3 Establish FAIR Digital Objects as citable exchange format for workflows for all EOSC services
O2.4 Establish FAIR Workflow Digital Objects as publishable scholarly objects

Task 2.1 Integration of EuroScienceGateway in EOSC (Aligned with objectives O2.1, O1.7, O5.2)

Particularly by improving the Findability and Accessibility of scientific workflows, this task focused on including EuroScienceGateway as a new service inside the EOSC ecosystem. Central to this is the further development of WorkflowHub (Gustafsson et al. 2025)—a registry for publishing and sharing computational workflows—advancing it from TRL-7 to TRL-9 to satisfy the operational standards expected of EOSC services. The WorkflowHub Club governance forum integrated needs and comments from developing user communities (WP5).

Following set best practices (M2.1, M2.2), a curated list[48] of high-quality workflows was compiled from actual use cases (WP5) and training materials (WP1). EuroScienceGateway (T3.4) made these workflows into executable FAIR Digital Objects by means of RO-Crates enriched with Persistent Identifiers (PIDs), hosted on WorkflowHub under the EuroScienceGateway Team[49].

Workflows and related research assets will be visible in EOSC catalogues and aggregators like OpenAIRE (T2.3, T2.4), therefore promoting their adoption by raising their discoverability and influence across scientific fields.

September 2022- August 2023 summary of achievements:

WorkflowHub was effectively registered as a data service in the EOSC Marketplace during the first reporting period with EGI's assistance, who helped define legal ownership and onboarding processes. WorkflowHub has been listed under ELIXIR UK since it is created by various ELIXIR-affiliated partners, with the Earlham Institute representing the legal entity.

This registration is a significant move toward developing WorkflowHub to TRL-9 maturity. Already covered are important areas including preservation policy and documentation of Persistent Identifiers (PIDs). Strengthening governance systems and guaranteeing great service availability will be the next phase emphasis, including using server failover systems.

We investigated more general EOSC integration. ELIXIR's Training Portal (TeSS) does this by means of current standards including Bioschemas, automatically combining WorkflowHub workflows and registering Galaxy Training Network (GTN) training resources.

Future intentions include investigating the viability of registering tools made available via usegalaxy.eu and evaluating more integration with other EOSC services including the EGI Workload Manager (DIRAC), ROHub, and the EGI Notebook. One major difficulty is allowing smooth user access to services—like starting Jupyter Notebooks—without demanding manual permissions via virtual organisations.

September 2023- August 2025 Achievements:

The EOSC landscape shifted during the second half of the project, as the EOSC Marketplace was decommissioned and transitioned to the EOSC EU Node, the first node in a federated EOSC ecosystem. This transition is ongoing, and WorkflowHub and Galaxy each have the potential to become part of the offering of either national nodes or a thematic node for life sciences. However, as both projects are spread across nations and domains, even in their core infrastructure, finding their exact “home” in the EOSC ecosystem is not straightforward.

Despite this uncertainty, work continued on enhancing capabilities for publishing workflow enhanced FDOs to EOSC. To support easier publication of data, workflows,

[48] https://workflowhub.eu/collections/13
[49] https://workflowhub.eu/projects/166

and workflow invocations to EOSC registries, Galaxy introduced an InvenioRDM export feature, which can publish datasets, histories, and workflow invocations as RO-Crates to repositories like Zenodo[50]. The rocrate-inveniordm package (Beer, Szente, and Chadwick 2024) also allows any RO-Crate (not just from Galaxy) to be uploaded to InvenioRDM instances, including Zenodo.

The FAIR Computational Workflows guidelines (Wilkinson et al. 2024) were developed by EuroScienceGateway joining the Workflows Community Initiative[51]. A book chapter is also in development based on this work. Finally, in mid-2025 a paper was published about WorkflowHub (Gustafsson et al. 2025).

Task 2.2 Reproducible and reusable FAIR Digital Objects
(Aligned with objectives O2.2, O2.3)

This task emphasized further development of the RO-Crate approach as the basis for handling workflows as first-class FAIR Digital Objects under the EOSC Interoperability Framework. The goal was to guarantee these objects are both reusable by various workflow services and interoperable across systems. Self-describing Digital Objects have extensible metadata schemas (profiles) that can be enriched by Workflow Management Systems (e.g., Galaxy) and other metadata pipelines (see Task 2.3).

Establishing community agreement on standard profiles for workflows and their execution runs, including rich provenance information and connections to outcomes, was a major component of the task. Building on prior work in EOSC-Life and CWLProv, the aim was to promote openness and, where feasible, complete reproducibility of computational studies (as described in D2.1[52] and evaluated with Task 3.4).

In cooperation with international organizations including the FAIR Digital Object Forum[53], the RDA Data Fabric Interest Group, and EOSC Core, the combination of RO-Crate with FAIR Signposting has been demonstrated to meet the formal FAIR Digital Object specification (see D2.1). WorkflowHub and the EuroScienceGateway platform show the practical application of this task.

In concert with the EOSC-01-05 project, we specified the best practices for persistent identifiers (PIDs) and minimum metadata criteria for workflow citation (related to Task 2.4).

September 2022- August 2023 summary of achievements:

Significant progress in formalizing and adopting the Workflow Run Crate profiles—now supported by at least eight workflow systems[54], including Galaxy—was made during the first reporting period. This task has resulted in the preparation of a publication for PLOS One (Leo et al. 2024) and cooperation with other workflow systems such Nextflow[55].

With this development highlighted at FDO2022 (de Geest et al. 2022) and the ELIXIR All Hands 2023 (Harrison et al. 2023), Galaxy now records workflow provenance and exports it as Workflow Run Crates (de Geest 2024). With cooperative effort scheduled during the ELIXIR Biohackathon[56], Galaxy will continue to evolve by capturing more thorough provenance and allowing RO-Crate import capabilities.

Working with FAIR-IMPACT[57], we have created PID procedures for Git repository-hosted workflows and included Software Heritage (SWHID) to continuously snapshot them.

[50] https://galaxyproject.org/news/2024-05-03-inveniordm-integration/ and https://galaxyproject.org/news/2025-03-10-inveniordm-integration-update/
[51] https://workflows.community/
[52] https://doi.org/10.5281/zenodo.13225792
[53] https://doi.org/10.52825/ocp.v5i.1273
[54] https://www.researchobject.org/workflow-run-crate/
[55] https://github.com/nextflow-io/nf-prov/releases/tag/1.4.0
[56] https://galaxyproject.org/news/2024-11-08-biohackathon-europe-2024-insights/ and https://elixir-europe.org/events/biohackathon-europe
[57] https://galaxyproject.org/news/2023-01-12-eurosciencegateway-wp2/

RO-Crate has been included in the ELIXIR 2024–2028 programme[58] under work packages concentrating on the environmental effect and scalable FAIR workflows; We have drafted a proposal for an ELIXIR Galaxy implementation study on FAIR tracks integration[59].

Finally, especially by means of its benchmarking activities in Task 3.3, where EuroScienceGateway is emphasized as a major partner, the next ELIXIR-STEERS project (beginning in 2024) has built on this task.

September 2023- August 2025 Achievements:

Improvements to the quality of Galaxy's exported Workflow Run RO-Crates by integrating the RO-Crate validation tool rocrate-validator[60] into Galaxy tests to ensure the specification is met.

Updates were made to the Galaxy web interface and the planemo tool to encourage workflow creators to follow the best practices for metadata collection, for example ensuring authors are represented using their full ORCID URI ("https://orcid.org/0000-0002-1825-0097" rather than "0000-0002-1825-0097"). This improves the quality of the knowledge graphs that this metadata is incorporated into (e.g. OpenAIRE), as when authors are represented in a consistent manner, it is easier to link together all their different work.

Other changes to the Galaxy codebase facilitated the capture of additional information about workflow steps and tools when exporting RO-Crates, and improvements to reproducibility when re-importing an RO-Crate back into Galaxy.

The development and stabilisation of Workflow Run RO-Crate support both as a first class input and output for WfExS-backend workflow engine orchestrator executions was performed along this period. Version 1.0.1 of WfExS-backend (published at the end of May 2025) currently supports both CWL and Nextflow workflows. It is able to export workflow execution provenance to a zip packed Workflow Run RO-Crate, as well as both reproducing and replicating the recorded analysis later using as starting point the previously generated Workflow Run RO-Crate. Last, but not the least important, WfExS-backend is also able to publish the generated WRROC in Zenodo, B2SHARE or Dataverse, among others, embedding the obtained DOI within the WRROC itself.

Finally, in June 2025, version 1.2 of RO-Crate[61] was released, formalising support for RO-Crate profiles in the core specification. This formalisation was heavily influenced by existing community use of profiles, including WorkflowHub use of Workflow RO-Crate and Galaxy and WfExS use of Workflow Run Crate[62]. An updated version (0.5) of Workflow Run Crate[63] was also published in June 2024.

Task 2.3 Using and enriching workflow FDOs (Aligned with objectives O2.1, O2.3, O5.3)

As Workflow FAIR Digital Objects (FDOs) exchange across EuroScienceGateway and EOSC services, this task emphasized improving their metadata richness. These workflow objects are meant to be shared, archived (e.g., in Zenodo), and registered in scholarly catalogues (e.g., OpenAIRE) and research knowledge graphs (e.g., DataCite PID Graph). Every service could have certain metadata needs and provide extra notes, therefore FDOs will get more and more FAIR-rich over time.

Our goal was to enhance the EuroScienceGateway FDOs using these metadata systems—especially knowledge graph services and metadata mining from academic publications. End users (WP5) benefited from this enrichment process in terms of

[58] https://elixir-europe.org/about-us/what-we-do/elixir-programme

[59] https://elixir-europe.org/internal-projects/commissioned-services/strengthen-data-management-galaxy

[60] https://github.com/crs4/rocrate-validator

[61] https://doi.org/10.5281/zenodo.13751027

[62] https://doi.org/10.1371/journal.pone.0309210

[63] https://www.researchobject.org/workflow-run-crate/profiles/0.5/workflow_run_crate/

discoverability, scientific annotation, and contextual direction. It lets us see things like appropriate deployment patterns, usage trends, and related workflows. In the end, this metadata enrichment guarantees compatibility with the larger EOSC ecosystem and helps improve workflow selection, more efficient Pulsar integration, and meta-scheduling (T4.3).

September 2022- August 2023 summary of achievements:

Officially starting in month 15, Task 2.3's preliminary work during this time frame established a strong basis for future enrichment of workflow FDOs utilizing metadata and knowledge graph technologies. By connecting WorkflowHub entries to related resources—including those accessible through the PID Graph—this work seeks to improve them.

One specific enrichment application is combining input dataset metadata. The FAIRtracks project backed by ELIXIR and the RDA has started early technical integration. This includes collaboration with the emerging RDA Working Group on "FAIRification of Genomic Annotations,"[64] supported by the EOSC-funded RDA TIGER initiative.

Technical first successes are:

- Allowing Galaxy to more effectively support interactive Docker-based tools.
- Galaxy's new "interactive client tool" category for simple integration.
- Including GTrack and GSuite as Galaxy datatypes for compatibility.
- Showing FAIRtracks integration at https://galaxy.fairtracks.net.

Active participation with the FDO Forum also helped to advance ideas. A preprint from a EuroScienceGateway beneficiary has generated debate on the architectural strategies to FDOs, therefore affecting how workflows might be machine-readable and interoperable. WorkflowHub put aspects of this idea into practice.

Working with FAIR-IMPACT, efforts were made to encourage several Dataverse-like repositories' adoption of HTTP Signposting, a simple FDO mechanism.

September 2023- August 2025 Achievements:

A Snakemake workflow (Hambley et al. 2025) was developed to build a knowledge graph based on RO-Crates downloaded from WorkflowHub. The workflow has five stages:

1. Scrape and store RO-Crate metadata from the WorkflowHub API
2. Merge all the RO-Crate metadata into a single RDF graph
3. Enrich the RDF graph with data from other sources (e.g. ORCID, Wikidata)
4. Consolidate and de-duplicate entities in the graph - e.g. where the same author is referenced in slightly different ways in different WorkflowHub records, these are combined into one entity
5. Publish the final results to Zenodo as a Workflow Run RO-Crate, including the final graph, intermediate outputs from each stage, and provenance metadata.

[64] https://www.rd-alliance.org/groups/fairification-genomic-annotations-wg/activity/

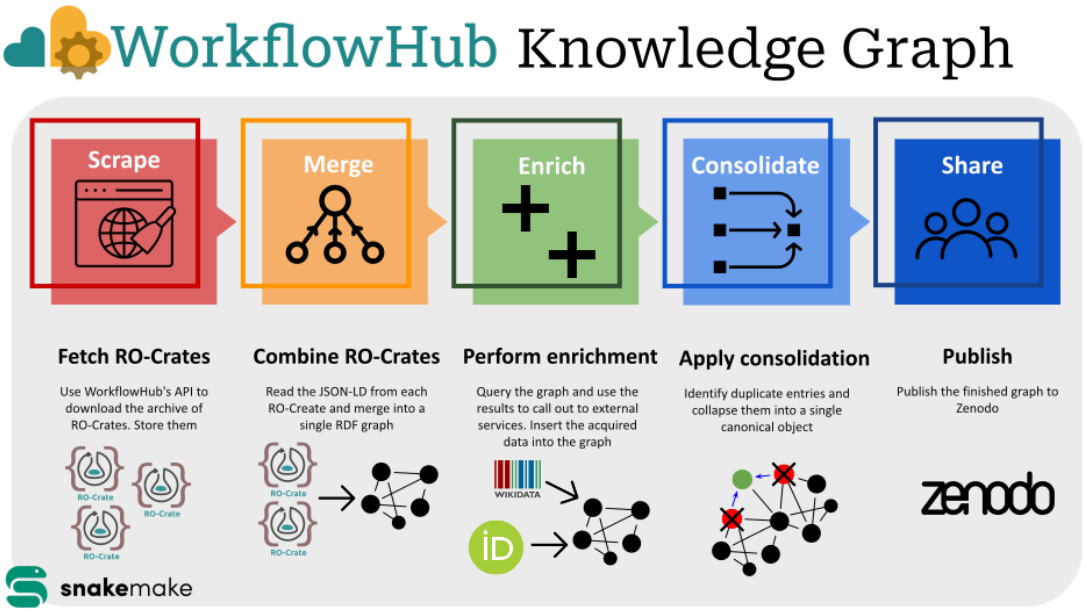

Figure 1. Steps in constructing the WorkflowHub Knowledge Graph.

An initial version of the knowledge graph was published in August 2024 (Hambley et al. 2024). A second version was published in August 2025 (Chadwick et al. 2025). The knowledge graph and the workflow used to create it are described in detail in D2.1[65] , and the source code is released on GitHub under the BSD 2-Clause license[66] .

The knowledge graph can be visualised using tools like Sampo-UI[67] and Zazuko Blueprint[68] . An example of a graph view of a single RO-Crate is shown in Figure 2.

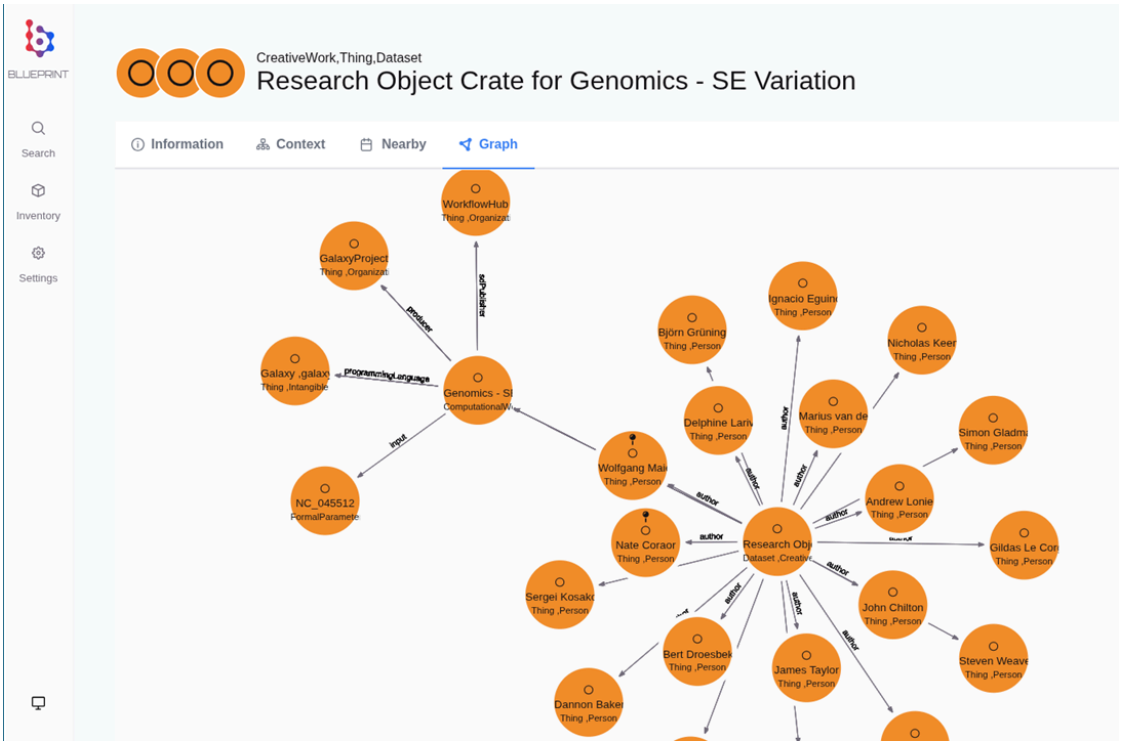

Figure 2. Graph view of a single RO-Crate, showing authors, programming language, publisher, and other workflow metadata.

The knowledge graph, being structured data, can also be used with LLMs and other ML tools, or queried directly using SPARQL. This could allow for the graph to be explored

[65] https://doi.org/10.5281/zenodo.16992674
[66] https://github.com/workflowhub-eu/workflowhub-graph
[67] https://seco.cs.aalto.fi/tools/sampo-ui/
[68] https://github.com/zazuko/blueprint

and queried using natural language, e.g. “fetch all the Galaxy workflows created by the person named <name>”. With further enrichment, this could expand to support queries spanning multiple data sources, e.g. “fetch all the Galaxy workflows that are associated with papers on the topic of astronomy.”

Task 2.4 FAIR workflows as scholarly objects in scientific publishing (Aligned with objectives O2.1, O2.3, O2.4, O5.3)

This task focused on placing FAIR workflows as essential, citable parts of the scientific publication system. Linking workflows and their metadata to research papers helps to improve the findability and reproducibility of scientific research and findings. Enriching workflow records with publication data therefore helps to raise their discoverability and reuse.

The aim was to create the technical tools to handle workflows as first-class academic items. This involves building strong connections between workflow FDOs and commonly used publishing platforms—from preprint servers (e.g. arXiv), rapid alerts (e.g. Astronomer’s Telegrams), and open-access repositories (e.g. OpenAIRE), to software-focused journals (e.g. JOSS) and conventional academic publishers.

A main goal was to establish WorkflowHub as a reliable registry authority for workflows, therefore allowing workflows to be referenced using persistent identifiers—DOIs from WorkflowHub, DockStore, or Zenodo—in accordance with current software citation policies (e.g. RDA FAIR4RS).

The task wanted to:

- Develop methods for drawing connections between publications and the workflows they describe or utilize.
- Encourage assessment of workflows and publications.

We worked together with relevant EOSC projects, such as INFRA-2021-EOSC-01-05, to create common criteria and recommendations for FAIR processes in academic communication.

September 2022- August 2023 summary of achievements:

This task has concentrated on creating workflows as academic outputs by means of cross-system direction, technical innovation, and cooperation with the larger research community.

A prototype tool[69] to extract workflow references from scientific papers and turn them into structured, verifiable RO-Crate objects (GitHub) was developed.

The team also looked at the difficulties of workflow publishing depending on external compute/storage infrastructure, especially for big data workflows in future large-scale astronomical observatories like SKA and CTA, stressing synergies between WP2 and WP5.

We worked on a project to specify a recuperative RO-Crate profile to assist long-term sustainability of published workflows; a journal article was prepared.

At last, MMODA's publishing processes were prototyped using automated technologies, therefore allowing the smooth export of workflows—for example, GW-backend—to LifeMonitor and WorkflowHub. We established a cross-project task group on workflow citation comprising publisher involvement and international software citation initiatives like RDA FAIR4RS.

September 2023- August 2025 Achievements:

Following best practices established in the other WP2 tasks (M2.1, M2.2), a curated list[70] of high-quality workflows was compiled from actual use cases (WP5) and training

[69] https://github.com/esg-epfl-apc/astro-paper-workflow-parsing swh:1:snp:8752ddd668010ad0be526f14d6ee0d7605e2f5c2

[70] https://workflowhub.eu/collections/13

materials (WP1). These workflows are hosted on WorkflowHub under the EuroScienceGateway Team[71], making them FAIR Digital Objects by default.

WorkflowHub bots were created to automatically register workflows from trusted high-quality sources with WorkflowHub. These sources include the Galaxy Training Network (GTN)[72] and the Intergalactic Workflow Commission (IWC)[73] from the Galaxy community, as well as nf-core[74], which curates bioinformatics workflows written in Nextflow (Langer et. al 2025). Both new workflows and updated versions of existing workflows are automatically registered.

WorkflowHub can register DataCite DOIs for workflows, using metadata provided by workflow creators. Once registered, the metadata is automatically incorporated into the DataCite Commons PID Graph and OpenAIRE Research Graph. As part of this integration, the "Workflow" resource type was added to DataCite.

Figure 3 shows the increasing adoption of workflow DOI registration through WorkflowHub, with significant year-on-year growth and a cumulative total of 639 workflow DOIs registered as of October 2025[75].

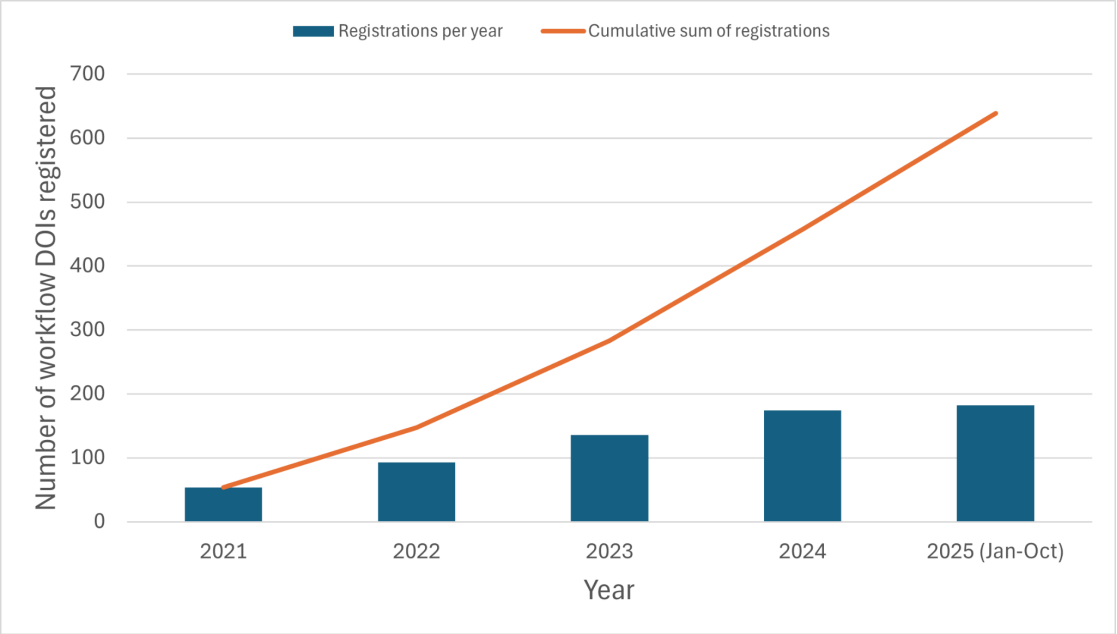

Figure 3. WorkflowHub DOI registrations from 2021 to 2025.

This also facilitates tracking of workflow citations. As of October 2025, there are 135 citations of WorkflowHub-generated DOIs in published works visible in Google Scholar[76].

WorkflowHub Publishers and Journals forum

Deliverables and Milestones
D2.1 Reproducible FAIR Digital Objects for workflows
D2.2 Publishing workflow enriched FDOs to EOSC
M2.1 Initial EuroScienceGateway Workflows registered
M2.2 EuroScienceGateway workflows registered as FDOs
M2.3 Initial EuroScienceGateway knowledge graph
M2.4 Integrated EuroScienceGateway knowledge graph

[71] https://workflowhub.eu/projects/166
[72] https://workflowhub.eu/projects/12
[73] https://workflowhub.eu/projects/33
[74] https://workflowhub.eu/projects/15
[75] https://commons.datacite.org/repositories/borg.00002-2021?resource-type=workflow
[76] https://scholar.google.co.uk/scholar?hl=en&as_sdt=0%2C5&q=10.48546%2Fworkflowhub.workflow&btnG=

1.2.3 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 https://usegalaxy.it.

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, the usage of tools like Ansible, Packer, and Terraform helped automated deployment and the compatibility with different platforms. The Open Infrastructure has been indeed tested with commercial cloud providers.

September 2022- August 2023 summary of achievements:

Every Pulsar endpoint has several key components: an HTCondor-based compute cluster, CERN-ViM File System (CVMFS) for tools 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 CNR, INFN, IISAS, CNRS, CESNET and EGI), 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 (pull request). 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 (pull request). The corresponding documentation is available here. 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 Figure 1. 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 online[77].

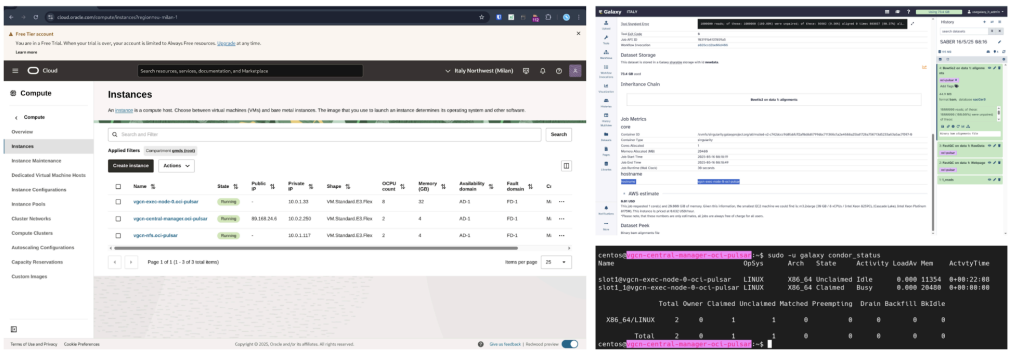

Figure 1. Pulsar deployment on Oracle Cloud

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 as shown in Figure 2. 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

[77] https://pulsar-network.readthedocs.io/en/latest/topics/oracle.html

developments include integration with Grafana dashboards and support for parallel workflow executions. Continuous monitoring is available at https://monitor.usegalaxy.it.

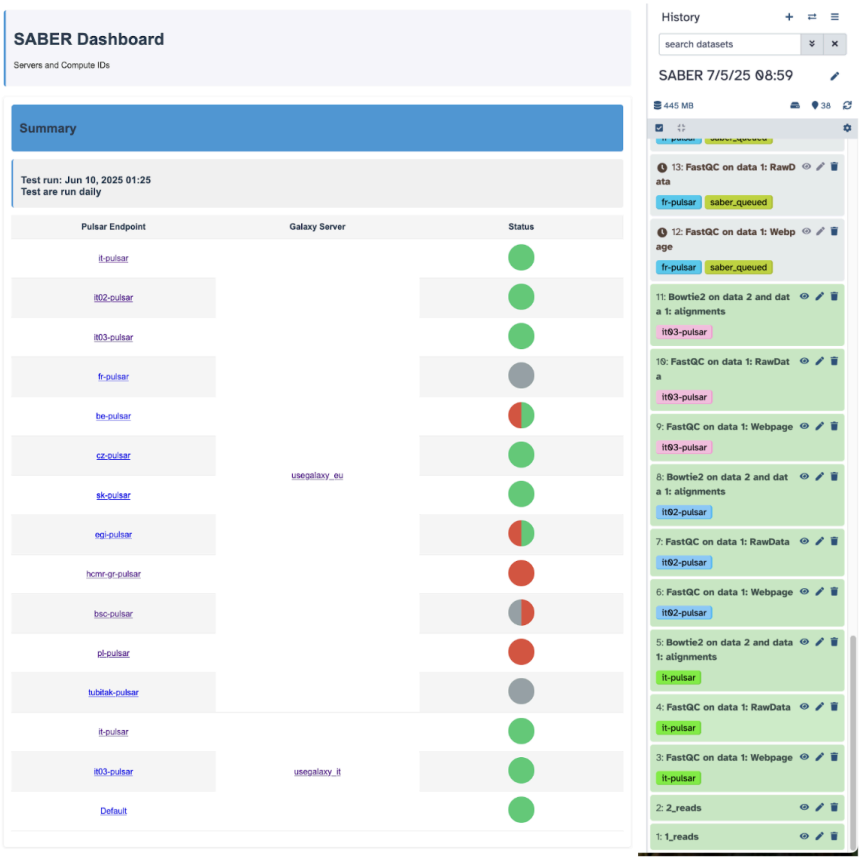

Figure 2. SABER report and Galaxy history

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[78] 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 would enable different workflow managers to send tasks to Pulsar endpoints, therefore promoting the 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, 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

[78] https://www.ga4gh.org/

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.

Every pulsar site is monitored via the public dashboard available at monitor.usegalaxy.it. With active participation from key infrastructure partners such as EGI, 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 13 endpoints, created by the consortium partners, using the pulsar deployment repository as depicted in Figure 3 and detailed in Table 1. Moreover, one endpoint has been added by the HCMR institute for the FairEase (Pull Request) project.

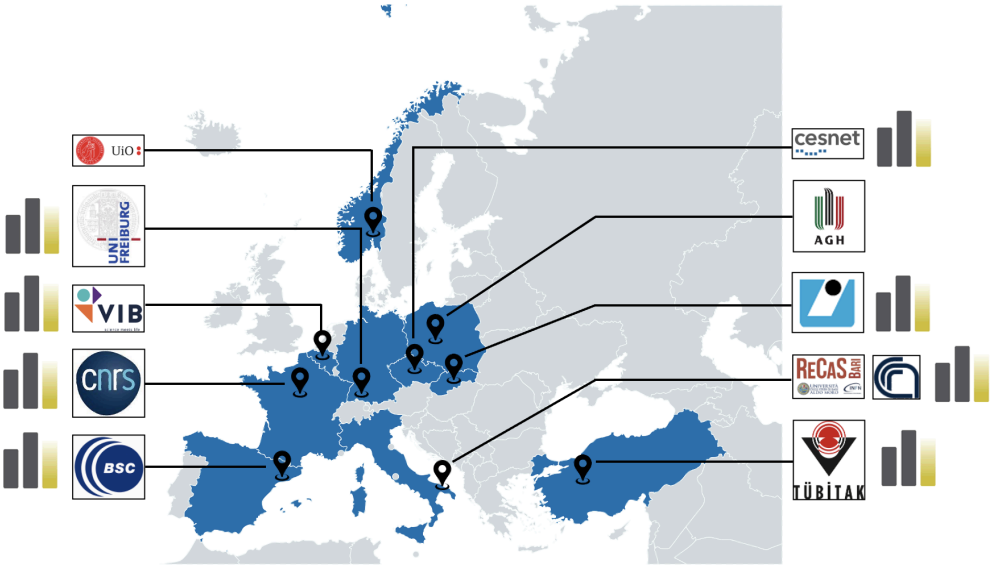

Figure 3. Pulsar endpoints map in EU

Table 1. Pulsar end points institutions, countries, and names

Institution	Country	Endpoint name
ALU-FR	Germany	DE01
CNR	Italy	IT01, IT02, IT03 and IT04
CNRS	France	FR01
IISAS	Slovakia	SK01
CESNET	Czech Republic	CZ01
VIB	Belgium	BE01
EGI and INFN	Italy	EGI01
BSC	Spain	BSC01
TUBITAK ULAKBIM	Turkey	TUBITAK01
Cyfronet	Poland	CFY01
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[79], a platform supporting the GA4GH TRS standard[80] 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

[79] https://wfexs-backend.readthedocs.io/en/latest/
[80] https://www.ga4gh.org/product/tool-registry-service-trs/

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.

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 as depicted in Figure 4.

WfExS-backend delegates workflow executions to the most appropriate supported workflow engine (currently either Nextflow or cwltool). Although Nextflow supports GA4GH TES to 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). The idea behind TESSAP is replacing the original docker client, maintaining both the parameter and behaviour compatibility as much as possible.

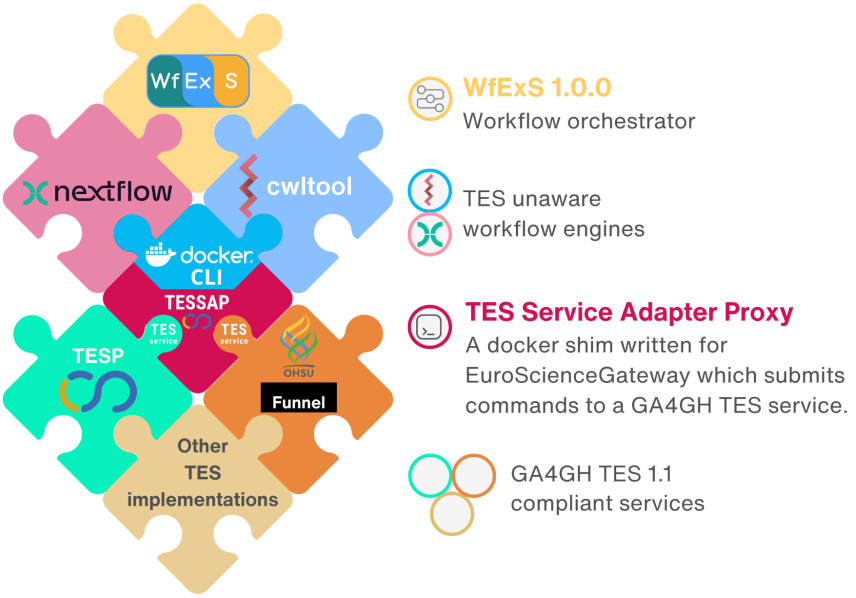

Figure 4. Modular design of TESSAP

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 as depicted in Figure 5.

Implemented	Subcommand	Nextflow	cwltool	WfExS-backend
✅	`docker run`	☑️	☑️	🔗
✅	`docker rm`	☑️	🔲	🔗
✅	`docker stop`	☑️	🔲	🔗
✅	`docker kill`	☑️	🔲	🔗

✅	`docker ps`	🔲	🔲	🔲
✅	`docker pull`	🔲	☑️	☑️
🎭	`docker stats`	🔲	☑️	🔗
✅	`docker inspect`	🔲	☑️	☑️
❌	`docker import`	🔲	☑️	🔗
🚧	`docker load`	🔲	☑️	☑️
❌	`docker build`	🔲	☑️	🔗
🚧	`docker save`	🔲	🔲	☑️
🚧	`docker images`	🔲	🔲	☑️
🚧	`docker tag`	🔲	🔲	☑️
🚧	`docker rmi`	🔲	🔲	☑️
🚧	`docker version`	🔲	🔲	☑️

Figure 5. Status of docker client commands implemented by TESSAP

So, most of previous subcommands from `docker` are already implemented (✅) or faked (🎭). Other ones are going to be implemented or faked (🚧), and a few ones are not going to be even tried (❌).
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 standardized 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:

1. 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 integrated, and a new user interface component was added to let users choose remote computer resources. To assist testing operations across work packages, a test replica of the EU production server was also established.
2. Using CINECA Cloud 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 TIaaS 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 jobs from running 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/7f0e51985536cfa9446b0d7e97555347a3c648c9
- 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, usegalaxy.be (https://usegalaxy.be), on the OpenStack cloud provided by the Flemish Super Computing Center (https://www.vscentrum.be/). During the project usegalaxy.be 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[81] 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, usegalaxy.fr. 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).

[81] https://gunicorn.org/

In addition to already deployed features (e.g. TIaaS, 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/).

usegalaxy.fr 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]: The Norwegian node of the Galaxy network, usegalaxy.no, carried out several infrastructure improvements in this period. Key activities included test deployment of a new AlmaLinux deployment, the restoration of the Pulsar endpoint, and improvements of the Cloud setup. The Squid proxy configuration was optimized by increasing the maximum file size to 90 GB per node (x2). The infrastructure, hosted in the Cloud environment, underwent maintenance operations related to hypervisor management to ensure long-term stability. Additionally, the team completed system upgrades to the latest Galaxy releases.

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 destination and Singularity images leveraging robust PBS job scheduler service of CESNET's metacentrum.cz which includes national ELIXIR CZ compute cluster and dedicated GPU machine for AlphaFold 2 jobs among other resources. On the Galaxy's side it uses Total Perspective Vortex (TPV) 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 tools.

For local indices and other data it accesses galaxy project's CVMFS. In terms of authentication it is connected to multiple AAI providers including LS-Login and e-infra.cz. We have configured TIaaS for training and support exporting and importing of RO-Crates including in combination with repository infrastructures of 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, usegalaxy.es, currently available as a development instance. The server is hosted and maintained by the Barcelona Supercomputing Center (BSC), with the domain usegalaxy.es 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 usegalaxy.eu. 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)

1.2.4 Work package 4: Building blocks for a sustainable operating model

Explain the work carried out in WP4 during the reporting period giving details of the work carried out by each beneficiary/linked third party involved.

Particularly during the COVID-19 epidemic, when it facilitated the quick processing of hundreds of thousands of SARS-CoV-2 datasets across Europe, the Pulsar Network has already shown its worth. Though successful, its present setup is not completely scalable; system administrators still need to manually intervene adding new storage or computing resources.

This WP developed the required improvements to important EuroScienceGateway components, specifically Galaxy and Pulsar, to overcome these challenges. The aim was to automate and simplify the integration of user-contributed storage infrastructure and computing. A smart scheduling system has been implemented to maximize the distribution of tasks over the available systems as the Pulsar Network grows with more resource providers, therefore guaranteeing efficient and balanced use of computing resources.

Online meetings: More than 30 meetings for WP4 have been held on a monthly basis over the lifespan of the EuroScienceGateway project, gathering partners from across all the relevant tasks. Monthly meetings have been key to track progress towards the project objectives and also triggered on-demand, topic-specific meetings to further advance the discussion and implementation of more complex tasks.

Contributions: Regarding Task 4.1 “Bring Your Own Compute”: INFN and ALU-FR have been working on the automated inclusion of Pulsar endpoints in Galaxy, UiO and ALU-FR have been improving the support in Galaxy and Pulsar for the ARC JobRunner, and EGI has automated the deployment of Pulsar and ARC using Infrastructure Manager. Regarding Task 4.2 “Bring Your Own Storage”: ALU-FR has been contributing to all the improvements in connecting Galaxy with external storage in the cloud, AGH/UST has added Onedata as an additional object store for Galaxy, and EGI has automated the deployment of MinIO using Infrastructure Manager. Regarding Task 4.3 “Implement a smart job-scheduling system across Europe”: VIB, ALU-FR and UiO have designed and implemented the TPV Broker, and CESNET has designed and implemented the Galaxy Job Radar with the collaboration of ALU-FR.

O4.1 Enable integration of computing resources from user allocations into the EuroScienceGateway
O4.2 Enable integration of existing user storage into the EuroScienceGateway
O4.3 Deliver a distributed caching mechanism across the Pulsar providers
O4.4 Implement a smart job-scheduling system to optimize the job distribution across Pulsar and user allocations.

Task 4.1 Bring Your Own Compute (BYOC) (Aligned with objective O4.1)

This task focused on allowing Galaxy users to directly include their own compute resources—such local Pulsar servers or personal cloud instances—into their Galaxy accounts. Users can register these resources and make them available for workflow execution by means of a secure configuration template. Coordinated by the smart meta-scheduler created in Task 4.3, the system identifies and uses these resources together with current infrastructure.

The task used automated deployment workflows created in T3.1 to assist users without technical knowledge to deploy Pulsar themselves but who have access to computing resources—e.g., cloud credits or HPC credentials. Galaxy users can start these

deployments straight from the Galaxy interface, therefore simplifying the BYOC process for more widespread use.

September 2022- August 2023 summary of achievements:

Initial demonstrations for connecting external compute resources to Galaxy were put in place during the first reporting period. Under the Galaxy preferences menu, a new user interface template was added allowing users to set Pulsar endpoints by defining information including message queue credentials and compute resource characteristics (e.g. RAM, CPU, GPU). In addition to this, another template lets people choose their desired Pulsar endpoint for job execution.

Work on integration also started to allow Galaxy to send jobs to ARC-based computing sites[82], which are an essential part of the Worldwide LHC Computing Grid (WLCG). Using the new ARC Python REST client, a prototype ARC job runner was created showing effective job submission and output retrieval from ARC resources. The execution automatically sends both outcomes and logs to Galaxy and comprises remote data staging and caching.

Galaxy was extended with WLCG IAM-based OIDC authentication to help secure access, enabling users to authenticate using tokens issued by supported identity providers. This feature was included into Galaxy's social-core authentication backend and employs a token refresher for ARC job submissions.

September 2023- August 2025 Achievements:

In the second reporting period: 1) Galaxy was integrated with EGI Check-in, an additional OIDC authentication proxy allowing new user communities to easily reuse their existing credentials; 2) scripts were developed to streamline the addition of new Pulsar endpoints to Galaxy, taking the information from the templates developed in the first reporting period; 3) the deployment and configuration of both Pulsar and ARC clusters were automated thanks to Infrastructure Manager; and 4) the Galaxy Job Runner for DIRAC was updated to improve its functionality. All details about these activities were reported in deliverable D4.1[83] Bring Your Own Infrastructure, and a summary is provided in the paragraphs below.

The integration of Galaxy with EGI Check-in brings eduGAIN[84] and other popular authentication sources (e.g. ORCiD[85]). EGI Check-in follows the AARC BluePrint Architecture[86] and implements existing AAI Interoperability Guidelines, making it compliant with the existing EOSC AAI Federation and ready to interoperate with the rest of the EOSC ecosystem. Since the integration with EGI Check-in in May 2024, 350+ users have logged in into the usegalaxy.eu instance. See Figure 6 below with a screenshot of the Galaxy login page with the EGI Check-in button.

[82] https://zenodo.org/records/15115715 and https://www.nordugrid.org/arc/arc7/
[83] https://doi.org/10.5281/zenodo.15729502
[84] eduGAIN: https://edugain.org/
[85] ORCiD: https://orcid.org/
[86] AARC BluePrint Architecture: https://aarc-community.org/architecture/

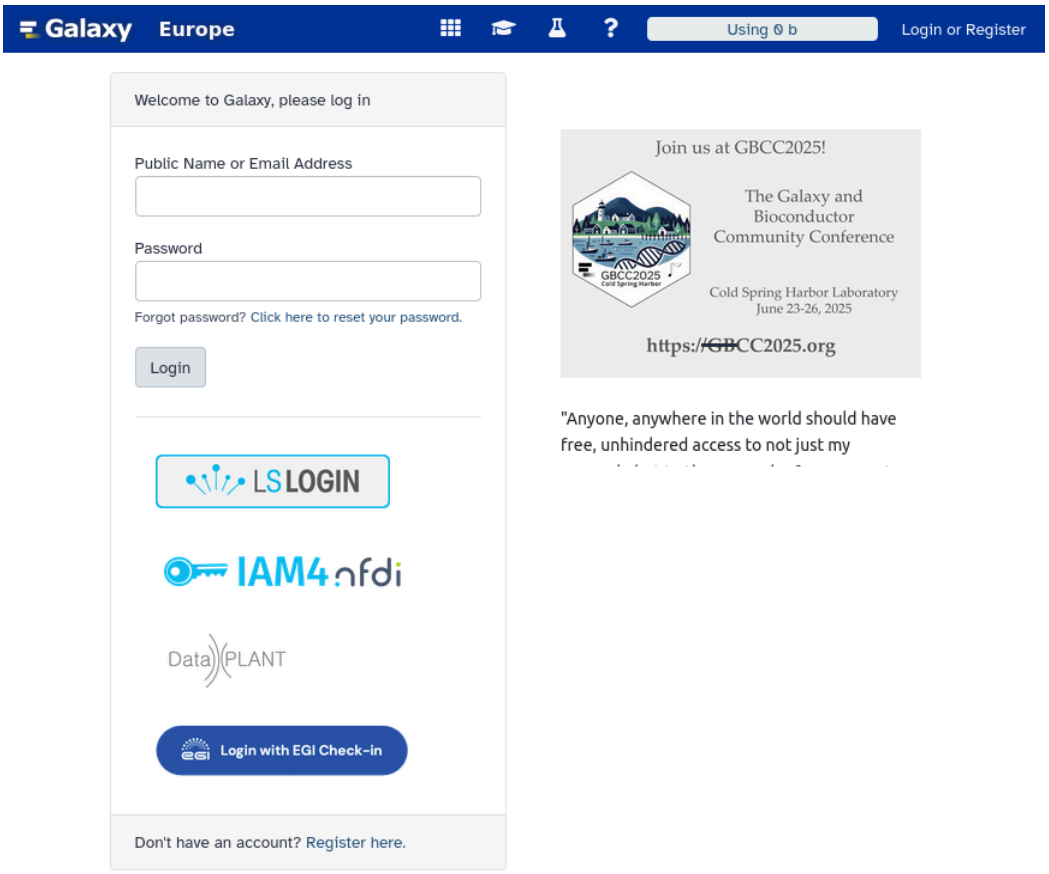

Figure 6. Screenshot showing the Galaxy login page with the EGI Check-in button.

The support for EGI Check-in opens the door to access services in the EGI catalog[87] as for example, the EGI Cloud Compute and Infrastructure Manager[88] for on-demand deployment of Pulsar and ARC clusters. Indeed, as part of this project, new TOSCA[89] templates and Ansible[90] playbooks have been created and added to Infrastructure Manager in a way that users with resources in the EGI Cloud Compute service can deploy and configure HTCondor[91] clusters with Pulsar and Slurm[92] clusters with ARC with the click of a button. See below Figure 7 with a screenshot of Infrastructure Manager selecting Pulsar for deployment.

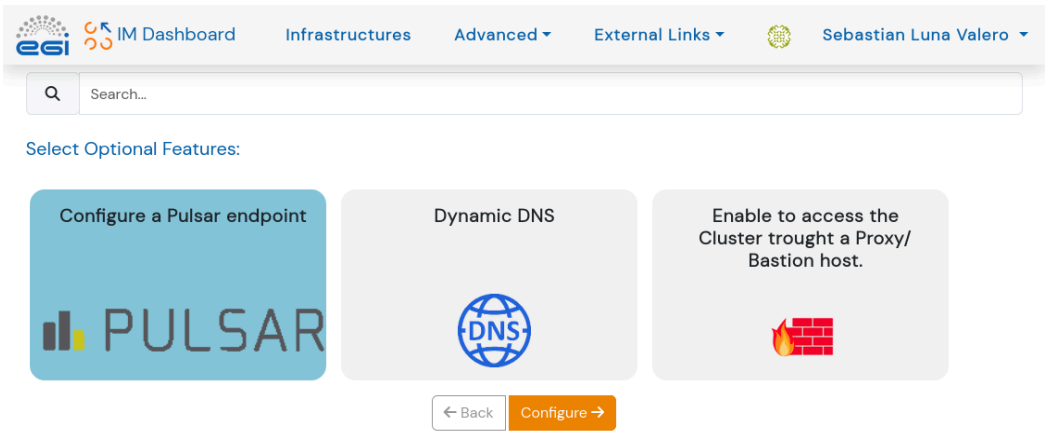

Figure 7. Screenshot of Infrastructure Manager with the Pulsar option selected.

[87] EGI Services for Research: https://www.egi.eu/services/research/
[88] Infrastructure Manager: https://im.egi.eu/
[89] OASIS TOSCA: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=tosca
[90] Ansible: https://www.ansible.com
[91] HTCondor: https://htcondor.org/
[92] Slurm: https://slurm.schedmd.com/documentation.html

A working version of the script to automate the inclusion of Pulsar endpoints in Galaxy is now available[93], which gathers the information submitted by the user via the form on the Galaxy instance, gives it the correct structure to create the pull request and submits it. The script, written in python, authenticates to the Galaxy instance (which the Pulsar endpoint should be attached to) via an API key (read from a locally stored secrets json file), and fetches the information submitted by the user, in json form, parsing it and storing it into variables, which are then reformatted in json/yaml suitable to integrate the infrastructure playbook repository with a pull request, which is then automatically generated and submitted. Secrets are safely stored using Ansible Vault. The script has a common "dry-run" option which can be used for testing and debugging purposes. Further developments are ongoing focusing on making the code (or part of it) more modular.

Since the last report, the ARC job runner is now in the final stages of being properly ported to work with new Galaxy and Pulsar functionality for remote job submission. The focus has been on the pieces needed to resolve the input, output and tool related paths in such a way that ARC itself can do the data and tool staging on the remote site. This is in order to handle generic Galaxy tools, and not just the prototype ARC specific tool that was described in section 3.2.1 of D4.1[94] . The pull request[95] have not yet been merged into the production branches of pulsar and galaxy. There is close communication with the maintainers of the Galaxy and Pulsar code bases, and the related pull requests will be merged into production once they have been reviewed.

Task 4.2 Bring Your Own Storage (BYOS) (Aligned with objectives O4.2, O4.3)

This task aimed to enable users to connect external storage resources to their Galaxy account, therefore transcending storage limits. Users are able to connect iRODS interface storage systems, Onedata repositories, or personal S3 buckets. Additionally, the job execution system has access to these user-provided storage endpoints, therefore enabling new access models like seamless EOSC-based storage integration into EuroScienceGateway.

The task requires creating a data locality mechanism. This system monitors the dataset's physical location or availability across Pulsar sites. The smart scheduler (created in T4.3) reuses the information to select geographically near compute resources to the data or those already cached, therefore optimizing data transfer and job efficiency.

September 2022- August 2023 summary of achievements:

Users are able to link their own storage resources to Galaxy, a long-standing objective of the core development team. A coordinated community effort brought this feature to life; it was then implemented using several combined pull requests detailed in the 1st periodic technical report.

The implementation provides a simple interface for setting up external object stores (e.g., S3, Azure), including support for visualizing storage characteristics including quotas, performance, location, and access controls. It also includes a templating tool for administrators to specify supported storage providers and necessary configuration parameters as well as safe storage of credentials[96].

A mapping system between datasets and storage endpoints has been created to assist with data locality, a fundamental need for efficient job scheduling under BYOS. Feeding into the job destination decision logic via Galaxy's Total Perspective Vortex (TPV), an early proof-of-concept tracks the geographic location of object stores and linked Pulsar compute nodes.

Onedata, a federated data management system providing consistent access to distributed storage has been added to the Galaxy as well. Based on a new lightweight pyfilesystem plugin (fs.onedatarestfs) and the onedatafilerestclient REST API, the Onedata team has created a Galaxy File Source Plugin. This lets Galaxy users directly access and use data kept in Onedata spaces.

[93] https://github.com/stefanonicotri/automatic_pulsar_endpoint_addition
[94] https://doi.org/10.5281/zenodo.15729502
[95] https://github.com/galaxyproject/galaxy/pull/20598
[96] https://github.com/galaxyproject/galaxy/pull/18127

September 2023- August 2025 Achievements:

In the second reporting period: 1) Galaxy has improved its support to connect with user-provided storage in the cloud; 2) the deployment and configuration of MinIO[97] was automated thanks to Infrastructure Manager; and 3) Onedata has increased the available options to connect Galaxy with external storage systems.

Building on the foundational work detailed in deliverable D4.1[98], the integration of user-managed external storage in Galaxy has reached a stable state. All core features described in D4.1, including support for user-defined object stores (BYOS) and file sources (BYOD), secure credential management via a vault, a graphical storage dashboard, and the ability to select storage at multiple levels are fully implemented and available in Galaxy.

Since D4.1, the following progress has been made, 1) the new features (object store and file sources) have been successfully deployed at UseGalaxy.eu and are actively utilized by more than 100 users, 2) object store[99] and file source templates[100] for major providers (AWS, Azure, GCP, and generic S3) have been further refined to ensure consistency and validation of required input fields, 3) initial documentation and training materials for end users and administrators have been developed to support adoption[101], and 4) user feedback is being collected from the live instance to inform iterative UI and UX improvements[102].

The steady progress in making it easier and safer for users to connect their cloud storage to Galaxy aligns with the goals of the EuroScienceGateway project. Work will continue to support more storage providers. For full technical details, see Deliverable D4.1.

Users with access to cloud resources in the EGI Federated Cloud can also benefit from a new TOSCA template in Infrastructure Manager, created in the context of this project, for the automated deployment of MinIO. This way, users can deploy a private MinIO endpoint and easily connect it to Galaxy. See Figure 8 below with a screenshot of Infrastructure Manager with the MinIO option selected. For full details please refer to Section 4.1.1 in deliverable D4.1 Bring Your Own Infrastructure.

[97] MinIO: https://min.io/
[98] https://doi.org/10.5281/zenodo.15729502
[99] https://github.com/galaxyproject/galaxy/tree/v25.0.0/lib/galaxy/objectstore/templates/examples and https://github.com/galaxyproject/galaxy/tree/v25.0.0/lib/galaxy/objectstore/templates/examples
[100] https://github.com/galaxyproject/galaxy/tree/v25.0.0/lib/galaxy/objectstore/templates/examples
[101] https://galaxyproject.org/eu/storage/, https://training.galaxyproject.org/training-material/faqs/galaxy/manage_your_galaxy_storage.html, and https://training.galaxyproject.org/training-material/faqs/galaxy/manage_your_repositories.html
[102] https://github.com/galaxyproject/galaxy/pull/19697 and https://github.com/galaxyproject/galaxy/pull/20070

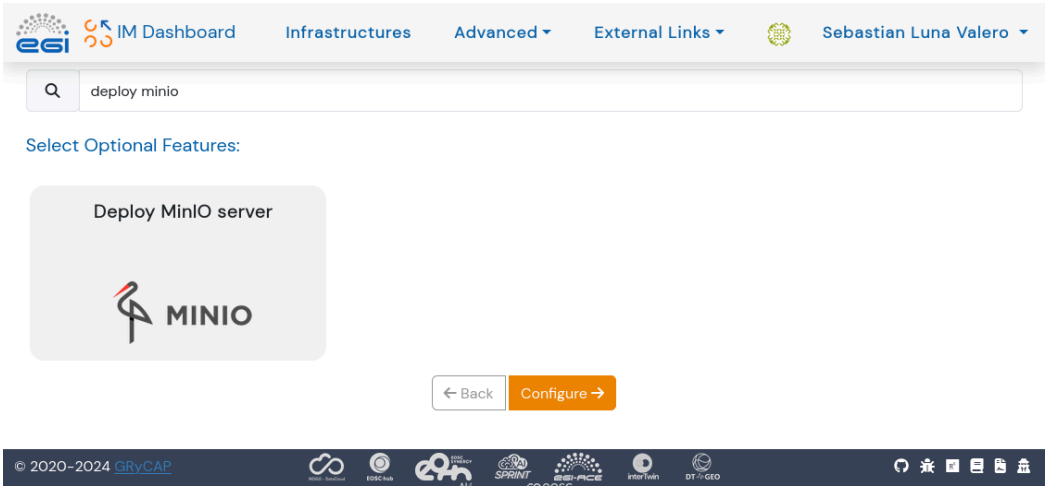

Figure 8. Screenshot of Infrastructure Manager with the MinIO option selected.

The Cyfronet team has continued work on integrating Onedata with Galaxy. The File Source Plugin, implemented in the first phase of the project, has been improved to better handle erroneous situations and misbehaving data providers, as described in section 4.2.1 of D4.1.

Another important effort was the implementation of Onedata Object Store. It enables Galaxy users to store their Galaxy data in Onedata Spaces, effectively extending their quota. Additionally, it offers a promising synergy with the smart job scheduler, as Onedata is equipped with data locality awareness. A detailed description is included in section 4.2.2 of D4.1.

To answer the requirements of Bring Your Own Storage (BYOS) and Bring Your Own Data (BYOD), adequate templates for Onedata have been added to Galaxy. Up to that point, server administrators had to set up Onedata[103] Remote File Sources and Object Stores. Thanks to the templates, it's now possible for Galaxy users to personally configure their data sources and Galaxy data storage locations, as depicted in section 4.2.3 of D4.1.

To simplify the onboarding of users that would like to use Galaxy with their data stored in Onedata, a series of comprehensive tutorials have been created and published in the official Galaxy Training Network[104]. They range from guides for beginners to more complicated tutorials for developers and those requiring advanced configuration.

To sum up, thanks to the efforts in EuroScienceGateway, Galaxy now features a comprehensive integration with Onedata. Both of these open-source systems are targeted at scientific communities representing different fields that make use of distributed data and computational analysis. This promises a better impact for both systems and a streamlined user experience.

All those improvements are live on the UseGalaxy.eu server and available in the general Galaxy project codebase, for any Galaxy portal to use.

Task 4.3 Implement a smart job-scheduling system across Europe (Aligned with objectives O4.4)

This task focused on creating a smart job scheduling system that best directs jobs sent to EuroScienceGateway to suitable compute resources, taking into account the user-provided computing and storage allocations brought in tasks 4.1 and 4.2.

The extended scheduler considers several factors, including data and compute locality (from Task 4.2), the current load, and available compute resources collected from Pulsar endpoints, as well as historical job data, which includes median queue times, run times,

[103] Onedata templates:
https://github.com/galaxyproject/galaxy/tree/v25.0.0/lib/galaxy/objectstore/templates/examples
[104] GTN Onedata tutorials:
https://training.galaxyproject.org/training-material/search2?query=onedata

and more.

By rescheduling jobs in the event of hardware problems, transient unavailability of a compute node, or job failures caused by lack of resources, this meta-scheduler not only recommend best execution sites depending on various strategies—e.g., runtime efficiency or carbon footprint—but also enhance general system reliability.

September 2022- August 2023 summary of achievements:

During the first reporting period the project started activities to extend the DIRAC[105] meta-scheduler and prepare it for integration as the smart scheduling solution for the EuroScienceGateway. One of the main activities involved creating a dedicated Virtual Organization (VO) called *vo.usegalaxy.eu*[106], which lets members of the EuroScienceGateway community access shared computational resources. Identifying and involving resource providers, writing a Service Level Agreement (SLA) to specify access conditions, and adding project partners into the VO was instrumental to get started.

The initial investigations to extend DIRAC to support the integration with Galaxy revealed that the amount of work to complete the task was higher than expected. On the one hand, Galaxy comes with Total Perspective Vortex (TPV)[107], a static meta-scheduler. Further investigations clarified that it would be more beneficial for the project to extend TPV and improve it (details reported below). This way, we would implement the desired smart scheduling in the EuroScienceGateway in a more efficient way.

The University of Oslo team also evaluated two adaptive meta-scheduling algorithms using the InterGridSim Simulator as a preparatory step:

- A fuzzy logic-based method combining locality-based job prioritization with resource/job need matching.
- A flexible approach using past resource performance data for decision-making.

Both models are meant to operate with centralized and distributed scheduling.

September 2023- August 2025 Achievements:

In the second reporting period: 1) an extension of Galaxy's TPV has been developed, the TPV Broker, to bring smart scheduling to the EuroScienceGateway; and 2) the Galaxy Job Radar has been created to visualize jobs floating across Europe as scheduled by the TPV Broker. More details about the two are provided in the paragraphs below.

The **TPV Broker**[108] is a new API endpoint that extends Galaxy's TPV. Galaxy uses TPV to schedule millions of jobs for hundred thousand users globally. While TPV has proven to be a robust meta-scheduling tool for Galaxy in the last years, there are areas of improvement that have been addressed in the EuroScienceGateway project, motivating the creation of the TPV Broker: 1) gathering live usage metrics from across the distributed computing endpoints connected to Galaxy in order to distribute the load across all sites. 2) adding latitude and longitude attributes to data stores and computing endpoints to allocate jobs as close as possible to the location of the data, and 3) addition of the meta-scheduling algorithms evaluated by the InterGridSim Simulator in the first reporting period.

The TPV Broker gathers live usage metrics from Pulsar endpoints thanks to producer/consumer scripts publicly available in the ansible-pulsar-utils repository[109]. Ansible playbooks are provided for the administrator of Pulsar endpoints to deploy the producer scripts. These scripts will query the batch job scheduler (i.e. HTCondor) and push usage metrics back to a central Influx Time Series Database. Another playbook is available for Galaxy administrators to deploy the consumer scripts that will query the

[105] DIRAC: http://diracgrid.org/
[106] VO in the EGI Operations Portal: https://operations-portal.egi.eu/vo/view/voname/vo.usegalaxy.eu
[107] Galaxy TPV: https://total-perspective-vortex.readthedocs.io/
[108] TPV Broker: https://github.com/usegalaxy-eu/tpv-broker
[109] ansible-pulsar-utils repository: https://github.com/usegalaxy-eu/ansible-pulsar-util

Influx Time Series Database to check usage across the Pulsar endpoints. Then, the meta-scheduling algorithms will assign jobs to particular endpoints with the aim to evenly distribute the load and reduce execution time. These and more details are included in Deliverable D4.2 and its accompanying publication[110]. The publication was accepted and presented at the 10th International Conference on ICT for Intelligent Systems (ICTIS[111]) in May 2025, and it will be published after a peer reviewed process in Springer Nature (DOI to be obtained).

Galaxy Job Radar is a web application and dashboard developed to visualize computational traffic across the Galaxy system. It provides various metrics, including the number of jobs across the network, the state of these jobs, and other relevant information, such as the location of computational nodes and the evaluation of scheduling within these nodes. All this information is presented on a world map (see Figure 9 below). Galaxy Job Radar also features a history replay of recent computations. Our vision is to integrate Galaxy Job Radar into Galaxy servers and to manage multiple Galaxy servers from a single instance as well. The project is still under development, with new features continually being added. More information about Galaxy Job Radar can be found in the Deliverable D4.2 document and in the associated publication.

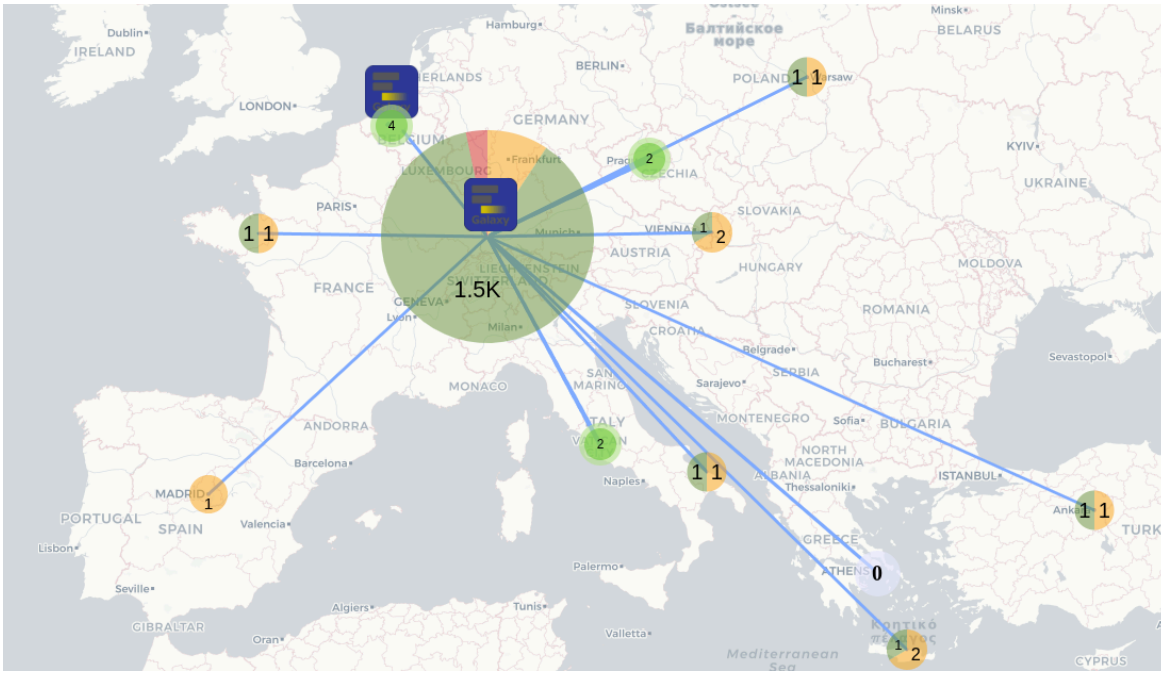

Figure 9. Screenshot of the Galaxy Job Radar in action.

The map in Figure 9 above shows the Galaxy instance running at the University of Freiburg, connected to pulsar endpoints across Europe: Belgium, Czech Republic, France, Italy, Poland, Slovakia, Spain, and Turkey. Pie charts in the map represent the amount of jobs in a specific site, and colors refer to the status of the job: running, queued and failed. This visualization of jobs assists in the validation of the meta-scheduling decisions made by the TPV Broker to ensure computational jobs are efficiently allocated.

Deliverables and Milestones
D4.1 Bring Your Own Infrastructure (compute, storage) Demonstrator (M24, 31st August 2024, Public)
D4.2 Publication on the smart job scheduler implementation (M30, 28th February 2025, Public)
M4.1 BYOC and BYOS integrated into ESG (M12, 31st August 2023)
M4.2 Meta-scheduler model for job optimisation available (M18, 29th February 2024)

1.2.5 Work package 5: Community engagement, adoption and onboarding

Explain the work carried out in WP5 during the reporting period giving details of the work

[110] https://doi.org/10.5281/zenodo.14936845
[111] ICTIS: https://ictisusa.com/

carried out by each beneficiary/linked third party involved.

Engaging with three different scientific communities—Biodiversity and Climate Science (with consequences for human health), Materials Science, and Astrophysics—WP5 focused on encouraging the adoption and practical use of the EuroScienceGateway services. From well-established use in Biodiversity/Climate to early-stage involvement in Astrophysics, these communities reflect a spectrum of Galaxy adoption maturity.

By highlighting success stories and use cases from these early adopters, WP5 aimed to show the influence of EuroScienceGateway, promoting broader community uptake and supporting Open Science best practices. WP5 also assisted the onboarding of new communities and helped customize the infrastructure to changing user needs in conjunction with WP1 (particularly Task 1.6).

WP5 promoted discussion among hardware, software, and application developers to preserve a community-driven approach by ensuring that user input—especially from researchers utilizing Galaxy for workflow composition and analysis—can guide the creation of technical components throughout the project.

O5.1 Customize EuroScienceGateway for domains by integrating relevant tools into Galaxy.
O5.2 Develop customizable pilot workflows to showcase results from the 3 early adopters.
O5.3 Onboard new communities to demonstrate the applicability of the new cross-cutting edge EuroScienceGateway services.

Contributions:
XXX
UKRI contributed to task 5.2

Task 5.1 Biodiversity and Climate Science (Aligned with objectives O5.1, O5.2, O5.3, O2.1, O2.3)

Through EuroScienceGateway, this task helped to integrate the biodiversity and climate research communities—both already familiar with Galaxy but historically operating in isolation—into a more unified and cooperative ecosystem.

Here, we had four main aims. First, working with climate and biodiversity researchers to find appropriate tools and datasets for EuroScienceGateway inclusion. Second, improving the accessibility, usability, and reuse of large, distributed, and authoritative datasets inside Galaxy. Third, assisting genome annotation at scale, especially to use data from the European Reference Genome Atlas (ERGA). Lastly, using both temporal (local time series) and spatial (geographical) viewpoints where feasible, developing adjustable workflows for investigating links between biodiversity indicators and climate change.

We demonstrated the advantages of co-designed workflows and underlined how EuroScienceGateway improves cooperation, resource sharing, and Open Science practices in these fields.

September 2022- August 2023 summary of achievements:

Biological diversity

During the project's first year, CNRS included and updated various genome annotation tools in Galaxy including Braker3, Helixer, Miniprot, Busco, Deepsig, and Assembly-stats. These tools were shared via project GitHub repositories and implemented according to Galaxy's development criteria. Testing and initial deployment took place on the French national Galaxy server (usegalaxy.fr) and are now being extended to other Galaxy sites worldwide.

Helixer, a GPU-supported artificial intelligence annotation tool, is one major development. It shows the advantages of combining with the Pulsar network created in

EuroScienceGateway, therefore allowing scalable access to GPU resources. These instruments taken together offer a strong basis for building best-practice workflows in genome annotation.

Especially inside the Vertebrate Genomes Project (VGP), these tools have been used on large-scale sequencing initiatives under the Earth BioGenome Project (EBP). A repeat masking workflow—a crucial stage in genome annotation—has been completed for submission to the Galaxy IWC repository as part of this effort. It is submitted also to Dockstore and WorkflowHub. Working with biodiversity research groups like ERGA and VGP helps to guarantee consistency with user requirements.

Climate Science

Rather than on single tools, the climate science group emphasizes creating comprehensive and reusable workflows. These workflows show how Galaxy, together with EuroScienceGateway, can enable complex data processing, compute-intensive analysis, and visualization for climate-related research.

Two workflows are being developed:

Originally created as a Jupyter notebook accompanying a 2021 Nature Communications publication, IceNet is a Galaxy implementation of the IceNet deep-learning pipeline for Arctic Sea Ice forecasting. The team developed several Galaxy tools for data concatenation, API requests to the Copernicus Climate Data Store, visualization, forecasting, and data preprocessing.

FArLiG (Forecasting ARctic LIchen browninG)

This workflow evaluates moss and lichen cover in northern Norway using high-resolution satellite images, meteorological reanalysis data, and machine learning. This pipeline is updated for more general use in vegetation monitoring and climate impact research depending on a notebook shown at EGU 2023. The team has prepared Galaxy tools to automate data download, preparation, and filtering.

These efforts taken together highlight the dedication of the EuroScienceGateway to provide significant and community-aligned tools and processes in both fields of climate science and biodiversity.

September 2023- August 2025 Achievements:

Biodiversity

In the first part of this project, we focused on the creation of high quality Galaxy tools for the annotation of genomes. During the second part of the project, two new quality control tools for genome annotation, OMark[112] and Compleasm[113], were also integrated as Galaxy tools.

Based on this collection of Galaxy tools, we focused during the 2023-2025 period on the creation of workflows and training material dedicated to the production of genome annotation data.

Specifically, 4 workflows were deposited to IWC: Repeat masking[114], Annotation with Maker[115], lncRNAs annotation[116], and Functional annotation[117]. Two other workflows are currently being deposited and should be available online by August 2025 (Genome annotation with Helixer[118], and with Braker3[119]). These workflows represent state-of-the-art and production-ready solutions for researchers willing to annotate newly

[112] https://github.com/galaxyproject/tools-iuc/tree/main/tools/omark
[113] https://github.com/galaxyproject/tools-iuc/tree/main/tools/compleasm
[114] https://workflowhub.eu/workflows/575
[115] https://workflowhub.eu/workflows/1323
[116] https://workflowhub.eu/workflows/1324
[117] https://workflowhub.eu/workflows/1262
[118] https://github.com/galaxyproject/iwc/pull/767
[119] https://github.com/galaxyproject/iwc/pull/768

sequenced genomes. They fit perfectly with assembly tools and workflows produced in the frame of the EBP (EarthBioGenome Project) and the VGP (Vertebrate Genome Project)[120].

To ease the adoption of this work, a collection of training material was created and deposited on the Galaxy Training Network: introduction to genome annotation slides[121], genome annotation with Helixer[122], genome annotation with Braker3[123], comparison of Braker3 and Helixer annotations[124].

Dissemination and onboarding was carried out by presenting the work achieved in various working groups (e.g. ELIXIR Biodiversity Community, EBP, VGP, ERGA), through the participation to training events (Galaxy Training Academy in 2024 and 2025, EBAII research school Roscoff 2024, Assembly and annotation session Rennes 2025), and talks and posters at national and international conferences (JOBIM 2024 and 2025, European Galaxy Days 2023).

Finally, the Galaxy Codex project emerged in 2025: its goal is to collect and curate all Galaxy resources available for specific scientific domains. A biodiversity Codex has been initiated, and a first curation effort was performed by the community at the ELIXIR All Hands Meeting in June 2025, in a workshop led by the Galaxy and Biodiversity ELIXIR Communities. The result of this effort will soon be a Biodiversity Galaxy Lab (in a specific homepage) with a curated and organised set of Galaxy resources (tools, workflows, training) for this community.

Climate

In order to render the climate software usable on different architectures several versions of the ESME (Earth Modelling System Environment) bundle for MPIch-4.2.3, OpenMPI-4.1.6/5.0.6/5.0.7 have been developed and are now available on Bioconda, see Figure 10. Some of the bundles which did not make it to Bioconda, like ParaStationMPI-5.10.0 and MVAPIch-4.0 have been deployed in a different channel. There was a request from the NorESM community who were looking for a solution to run their model with containers, and we have been involved in an Extended User Support for Norwegian Research Infrastructure Services (NRIS). This was a great opportunity to test ESME for a genuine use case in real conditions, and compare the performance with the bare-metal counterpart optimized for Intel compilers & IMPI on Betzy. So far the results exceeded our expectations, with significant speedup for their reference case, but also faster compilation and init times, and bitwise reproducibility on different hosts like Fram (Intel Xeon CPUs) and Betzy (AMD “Rome” CPUs). Recently the tests have been moved to the Lumi HPC (AMD “Milan” CPUs).

[120] https://workflowhub.eu/search?q=vgp#workflows

[121] https://training.galaxyproject.org/training-material/topics/genome-annotation/tutorials/introduction/slides.html

[122] https://training.galaxyproject.org/training-material/topics/genome-annotation/tutorials/helixer/tutorial.html

[123] https://training.galaxyproject.org/training-material/topics/genome-annotation/tutorials/braker3/tutorial.html

[124] https://training.galaxyproject.org/training-material/topics/genome-annotation/tutorials/comparison-braker-helixer-annotation/tutorial.html

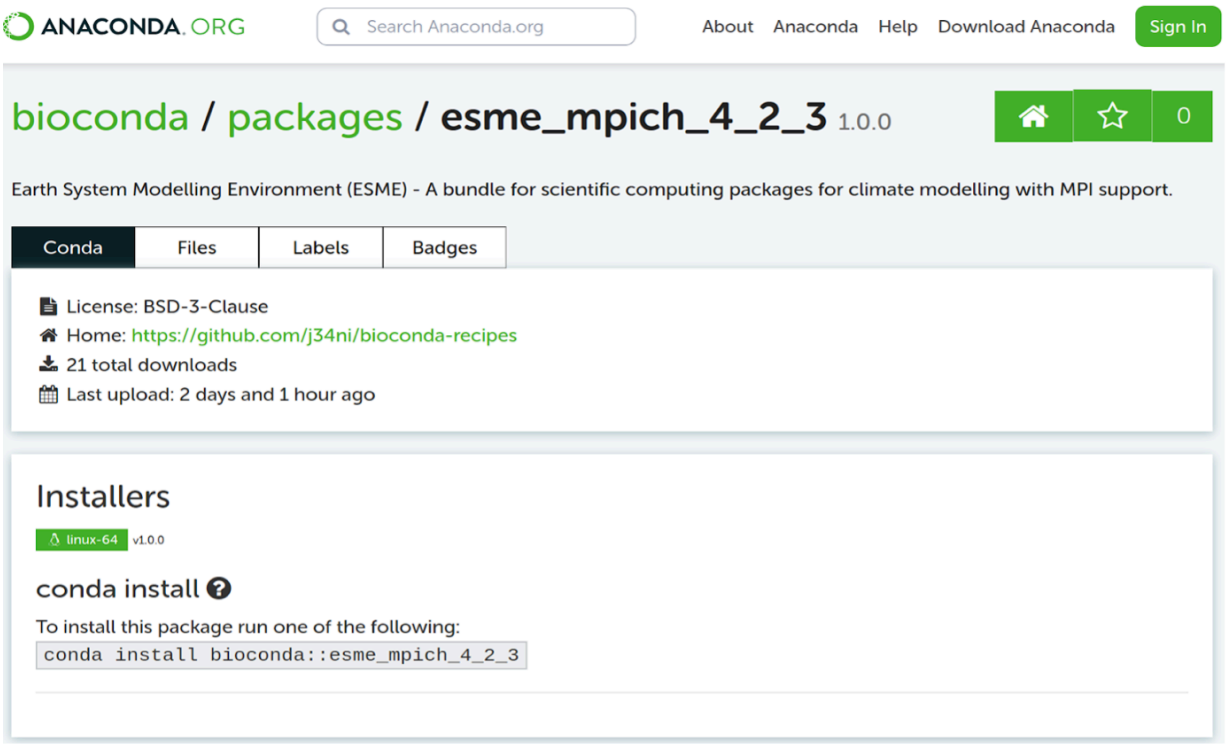

Figure 10. Earth System Modelling Environment (ESME) package on conda

A newly developed JupyterGIS interactive tool has been deployed to Galaxy Europe[125] (see Figure 11 below) and also on OpenOnDemand (a similar version for National e-Infrastructure for Research Data - NIRD) where it is still work-in-progress.

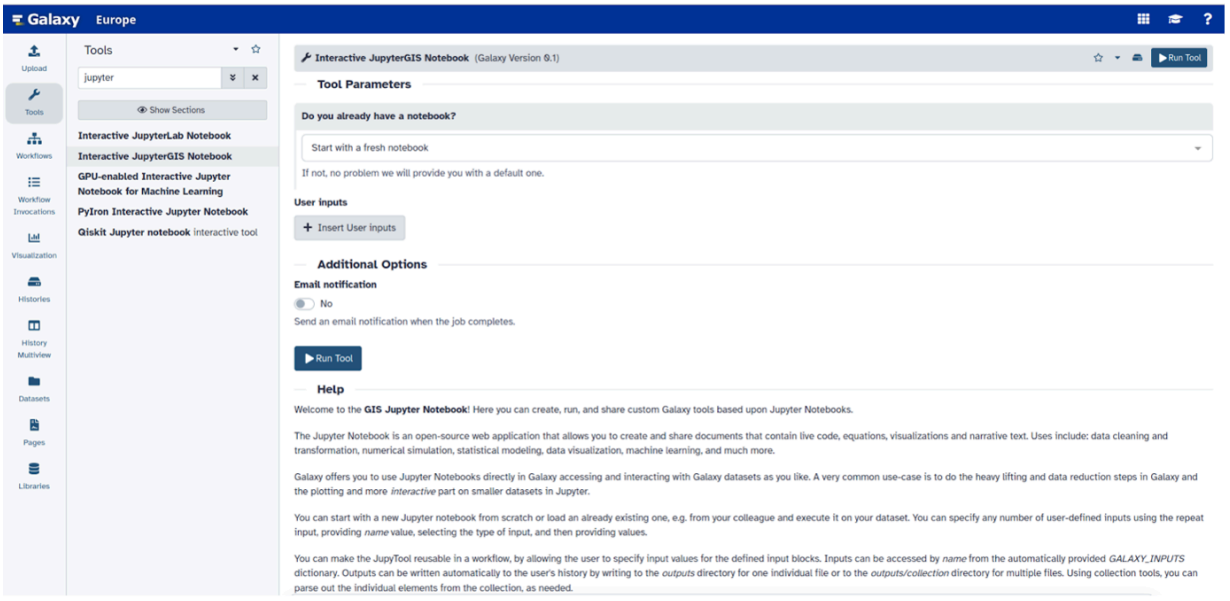

Figure 11. Interactive JupyterGIS Notebook available on usegalaxy.eu

Task 5.2 Materials Science (Aligned with objectives O5.1, O5.2, O5.3, O2.1, O2.3)

The Muon Spectroscopy Computational Project (MSCP) develops mathematical models and sustainable software tools for the interpretation of muon spectroscopy experiments. Galaxy was chosen to facilitate the easy access to these tools by the muon community, and also to encourage the development of a FAIR (Findable, Accessible, Interoperable,

[125] https://galaxyproject.org/news/2025-05-20-jupytergis/ and https://training.galaxyproject.org/training-material/topics/climate/tutorials/jupytergis_collaboration/tutorial.html

Reusable) methodology for performing computer simulations of muon experiments.

September 2022- August 2023 summary of achievements:

Initially (in the first year of the project), this task comprised the integration of the existing MSCP software pymuon-suite[126] and muspinsim[127] into Galaxy following best practices. Following a period of iterative development, this resulted in 11 tools, all of which are listed on the Tool Shed[128]. They are available on a dedicated Galaxy instance[129] maintained by the Scientific Computing Department of the Science and Technologies Facilities Council, in the UK, to promote FAIR practices within the muon science community. The Galaxy tools deployed in this instance were also deployed in the European Galaxy instance, to improve collaboration and make use of EuroScienceGateway services' features. Community onboarding was another key activity. The team collaborated with the international muon science community—including users of national and international muon sources—to grasp their needs and optimize the use of Galaxy services. And then the team used its Galaxy expertise to onboard other members of the large-scale experiments community.

Furthermore, we presented the work done on materials science, astrophysics, climate science and genomics at the PASC23[130] conference, which confirmed the potential of Galaxy as a robust and flexible workflow management platform, and the team began to use its expertise to develop Galaxy tools in other fields of materials science. In particular, we started to develop a set of prototype Galaxy tools to analyse X-ray Absorption Spectroscopy (XAS) data from catalysis experiments. The first prototype of these tools was presented at the UK Catalysis Hub Summer Conference 2023[131].

September 2023- August 2025 Achievements:

During the 2nd periodic report, the software developed by the MSCP consolidated as part of the standard software tools[132] used in muon science. We gave a tutorial[133] on the pymuon-suite Galaxy tool at the ISIS muon training school 2024[134] and have worked on simulating muon experiments applied in catalysis[135]. The workflow corresponding to this tutorial is also available on WorkflowHub[136] and provides a well defined and accessible solution to the problem of finding the stopping site of an implanted muon when performing muon spectroscopy. A new tool, mudirac[137], was added in July 2024. This software is of particular value to the community as one of the few packages that simulates negative muon experiments.

In this period, we also finalised 7 tools for XAS wrapping the Larch[138] Python library. As a demonstration of how these Galaxy tools can provide a FAIR method for analysing XAS experimental results, we repeated analysis described by 9 papers on different catalysis experiments[139]. The 9 resulting Galaxy workflows corresponding to each paper are available on WorkflowHub, and the export RO-Crate feature of Galaxy was also used to produce a single digital object with all datafiles, parameters, versions of software, and provenance between steps. This approach demonstrated how Galaxy can be used in this community to improve reproducibility and reduce the burden on researchers to produce

[126] https://github.com/muon-spectroscopy-computational-project/pymuon-suite
[127] https://github.com/muon-spectroscopy-computational-project/muspinsim
[128] https://toolshed.g2.bx.psu.edu/repository/browse_repositories?operation=repositories_by_category&id=9dca5b8f68c509af
[129] https://materialsgalaxy.stfc.ac.uk/
[130] https://pasc23.pasc-conference.org/presentation/?id=msa156&sess=sess169
[131] https://ukcatalysishub.co.uk/event/watch-now-uk-catalysis-hub-summer-conference-2023
[132] https://iopscience.iop.org/article/10.1088/2516-1075/adcb7c
[133] https://gxy.io/GTN:T00402
[134] https://indico.stfc.ac.uk/event/954/
[135] https://pubs.acs.org/doi/10.1021/acs.inorgchem.4c05126
[136] https://workflowhub.eu/workflows/1586
[137] https://toolshed.g2.bx.psu.edu/repository/browse_repositories_by_user?sort=name&operation=view_or_manage_repository&id=487662fd062ee564
[138] https://xraypy.github.io/xraylarch/
[139] https://doi.org/10.1002/cctc.202401676

high quality supporting data to their publications (in contrast to existing approaches which rely on local, interactive analysis where the details of intermediary steps are lost or not documented). We presented our results from this work at the 2024 Galaxy conference[140] and at the 2024 NoBugs[141] event, where we also organized a workshop[142] on applications of Galaxy for materials science.

Finally, we continue to expand the Galaxy approach to new areas of materials science, and are now working with the Extreme Photonics Applications Centre[143] (EPAC), at the Science and Technology Facilities Council, using Galaxy to implement an end-to-end simulation system for EPAC. Simulation of experimental conditions using virtual representations of instruments ("digital twinning") is also of interest at other STFC supported facilities, and internationally. There are multiple stages between the instrument settings that a beamline scientist can control and the final experimental output, and a workflow is a natural way of representing this which allows the effect of settings changes to be propagated with ease. Furthermore, the ability to define tools independently of each other is useful as often there are different simulation methods available for specific stages that users may want to easily swap between. Ultimately, the end goal is to make experiments more efficient by identifying desired conditions in advance so settings do not have to be tuned as much during the experiment. We presented our results on this project at the 2025 Galaxy and Bioconductor Conference[144].

Task 5.3 Astrophysics (Aligned with objectives O5.1, O5.2, O5.3, O2.1, O2.3)

This task focused on introducing Galaxy to the astrophysics community; its aim was to include often used data sources and analytical tools into the EuroScienceGateway platform. Working with Work Package 2, we integrated FAIR ideas into the practices and processes of astroparticle physicists, cosmologists, and astronomers.

First, with support for large-scale computation via the Pulsar Network, we identified and used appropriate tools employed in telescope data analysis and astronomical research on a dedicated Galaxy instance customized for astrophysics. Besides, we aimed to facilitate smooth access to astronomical datasets inside Galaxy (in line with Task 4.2), therefore supporting integration with outside data sources. In addition to this, we showed EuroScienceGateway's benefits for processing distributed, large-scale datasets in "multi-instrument" and "multi-messenger" astronomical research settings. We also created Galaxy workflows for astrophysical data analysis and simulations to show Galaxy's ability to manage large data in this field. To improve accessibility and user involvement, we included Galaxy API functionalities to let external astronomical data services start processes directly. Lastly, in alignment with WP2, we promoted FAIR workflows and data reuse, supporting reproducibility and openness in astrophysics data analysis.

September 2022- August 2023 summary of achievements:

Efforts on Galaxy's introduction of the astrophysics field concentrated on facilitating access to necessary astronomical data formats and tools during the first periodic report. Astroteam, a committed team, was established in the Galaxy ToolShed to provide the basis for hosting and sustaining astrophysics-related tools[145]. The inclusion of core support for the FITS (Flexible Image Transport System) format, the standard in astronomy for storing multidimensional sky images and data tables, was one of the first significant advances[146]. This improvement lets all significant Galaxy instances identify FITS files and show fundamental metadata, therefore enabling their more general use across the platform.

140 https://docs.google.com/presentation/d/1-t_f95BVq31klQCwBeAlUe0ljHY8PEmuJImhdQQAz4c/edit#slide=id.p

141 https://indico.esrf.fr/event/114/contributions/775/

142 https://indico.esrf.fr/event/144/

143 https://www.clf.stfc.ac.uk/Pages/EPAC-introduction-page.aspx

144 https://gbcc2025.bioconductor.org/program/scientific_program/

145 https://toolshed.g2.bx.psu.edu/repository?repository_id=f3d18dedd57fa589&changeset_revision=ea39f416af9d

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

A visualization plugin based on AladinLite was created and added to Galaxy to go along with this, therefore allowing users to interactively investigate FITS sky pictures and contrast them with publicly accessible astronomical catalogs. A new Galaxy tool was presented allowing users to browse and get data straight from IVOA-compliant archives, addressing another major issue of astronomers: access to large-scale astronomical archives[147]. This effort greatly enhanced Galaxy's compatibility with current astrophysical processes.

To allow the manipulation and conversion of FITS files into tabular formats compatible with Galaxy, additional tools based on the commonly used Astropy Python package were created[148], therefore showing beneficial synergies with current tools from other scientific fields. To enable more cooperation and adoption, all published and prototyped tools are kept in a public repository[149]. Moreover, we included formerly created astronomical tools from the MMODA project and RenkuLab among other platforms[150]. Paving the way for analyses using data from a wide range of optical, X-ray, gamma-ray, neutrino, and gravitational wave observatories—including DESI, ANTARES, LIGO/Virgo, Fermi/LAT, HESS, Euclid, Gaia, and INTEGRAL—a prototype converter[151] was built to turn these tools into Galaxy-compatible workflows.

We outreached to involve the larger astronomical community once the sufficient degree of functionality had been reached. A presentation at the Swiss SKA Days 2023[152], where we highlighted the tools and capabilities created thus far and gathered insightful comments for next enhancements, was one of the main events.

September 2023- August 2025 Achievements:

In the 2nd half of the project, we focused on dissemination, communication, and sustainability of the assets developed within T5.3.

Following through with meeting the astro community where it is - adopting community workflows and tools developed in jupyter notebooks, - we refined our notebook-to-galaxy conversion tool[153], and published a public guide[154] detailing this process.

Largely following this process, we developed (and supported the community in developing) more than 30 domain-specific astronomical tools, enabling a diverse range of astronomical use cases. The tool collection covers allows to analyse data of telescopes from radio (MWA, MeerKAT), to optical (Euclid, DESI, etc), to X-ray (eROSITA, INTEGRAL), to Very High Energy Gamma-ray (HESS, MAGIC, LST1).
These tools are published on astro-galaxy staging instance[155] following declarative gitops-like integration process (see public repository[156] serving as a reference and development space for the staging instance tool collection).

Another very significant collection of astronomical tools for Galaxy has been derived from NASA's HEASoft integrated package. It includes hundreds of data analysis routines for reducing data of most NASA space telescopes.

---

147 https://galaxyproject.org/news/2023-09-07-esg-wp5-astronomy-archives/

148 http://toolshed.g2.bx.psu.edu/repos/astroteam/astropy_fits2bitmap/astropy_fits2bitmap/0.3.0+galaxy0, https://astronomy.usegalaxy.eu/root?tool_id=toolshed.g2.bx.psu.edu/repos/astroteam/astropy_fits2csv/astropy_fits2csv/0.2.0+galaxy2, and https://astronomy.usegalaxy.eu/root?tool_id=toolshed.g2.bx.psu.edu/repos/astroteam/astropy_fitsinfo/astropy_fitsinfo/0.2.0+galaxy2

149 https://github.com/esg-epfl-apc/tools-astro/tree/main

150 https://galaxyproject.org/news/2025-06-11-astro-text-analysis/

151 https://github.com/esg-epfl-apc/tools-astro

152 https://galaxyproject.org/events/2023-09-06/

153 https://galaxyproject.org/news/2025-06-16-jupyter-to-tool/ and https://galaxyproject.org/news/2025-06-16-jupyter-to-tool/

154 https://odahub.io/docs/guide-galaxy/

155 https://galaxy.odahub.fr/

156 https://github.com/esg-epfl-apc/tools-astro/

Selected tools are published on the EU Galaxy instance, : https://astronomy.usegalaxy.eu/, established the EU Galaxy entrypoint for the astronomical community.

Using some of the developed tools we published several workflows in WorkflowHub[157].

We put an emphasis on the workflows linking different stages of astronomical knowledge generation, specifically one of the workflows starts with astronomical micropublications (ATels) and finishes with publication-ready publication[158]. This sort of end-to-end process is crucial for transient source analysis, when time of reaction is crucial to avoid missing important cosmic events. Early version of this tool presented AstroCOLIBRI broker and transient analysis workshop, and has since been published on WorkflowHub[159].

T5.3 triggered several spin-off activities, incorporating Galaxy in different contexts of astronomical research.

OSCARS FAIR imaging project explores multi-disciplinary exploitation of imaging tools within the Galaxy platform. The project kick-off was featured in several press communications[160].

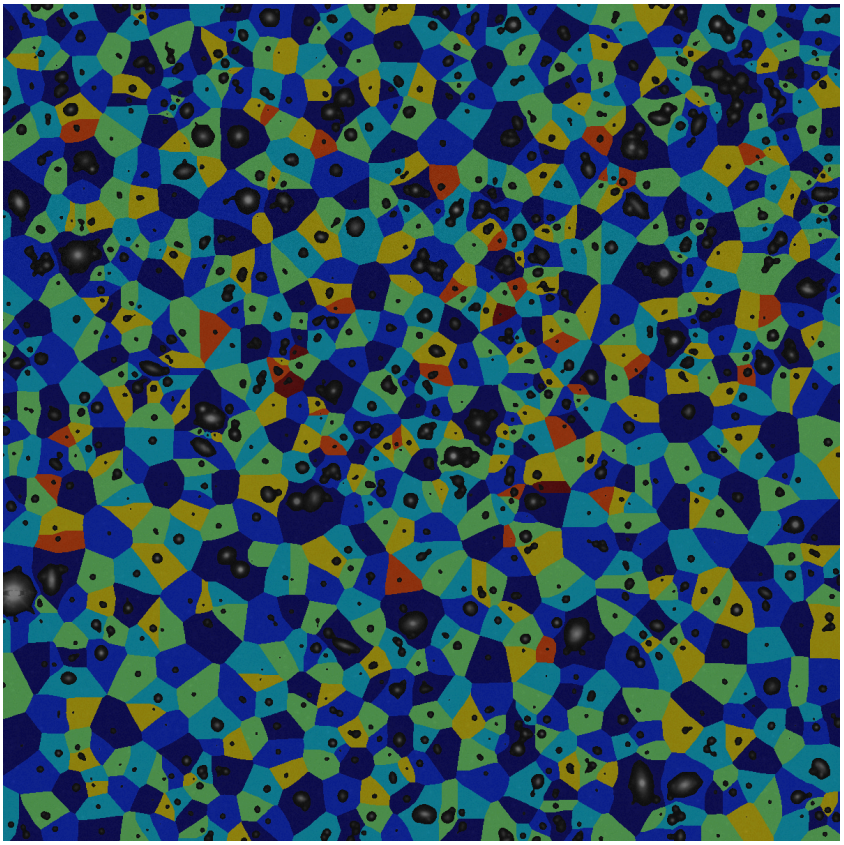

Figure 12. A dense field of galaxies analysed with Galaxy tools. See tutorial[161] for more details.

Thanks to the communication, dissemination, and interoperability development effort of T5.3, Galaxy had become an integral part of the Swiss AstroORDAS project[162] as well as the Swiss ORD Solidipes[163] project.

[157] https://workflowhub.eu/search?q=astronomy#workflows
[158] https://workflowhub.eu/workflows/1353
[159] https://workflowhub.eu/workflows/1353
[160] https://oscars-project.eu/projects/fair-image-analysis-across-sciences
[161] https://training.galaxyproject.org/training-material/topics/imaging/tutorials/voronoi-segmentation/tutorial.html
[162] https://www.astro.unige.ch/astroordas/
[163] https://gitlab.com/solidipes/solidipes/

Galaxy is represented in ACME Astrophysics Centre for Multimessenger studies in Europe – ACME - a multi-year project aiming to strengthen ties and enhance interoperability in the EU Astro community. ACME kick-off was held in September 2025 in Paris[164]. Positioning Galaxy in this project guarantees long-term visibility of T5.3 assets.

Finally, Galaxy was introduced as one of the possible science analysis platforms within next-generation astronomical observatories: CTAO[165] and SKAO[166].

Task 5.4 Mentoring and onboarding new communities (Aligned with objectives O5.3, O1.6, O1.7)

In cooperation with WP1, we will organize community-specific outreach events to connect with local communities and collaborate with other European initiatives that offer Galaxy training/mentoring efforts in EOSC (Skills4EOSC, AquaInfra, FAIR-EASE, BY-COVID, EOSC4Cancer and other), EuroHPC, EGI, ELIXIR, EuroBioImaging and new communities that show interest to join the EuroScienceGateway and want to make their own workflows or infrastructure accessible. We will also address cross-domain community needs to favor the development and the deployment of multi-disciplinary workflows to the EuroScienceGateway.

The task members will work together in documenting their domain-driven experience in joining the Galaxy community, sharing experiences and lessons learned, and abstracting away common strategies. These outcomes will be made available on the Galaxy Community Propeller onboarding cookbook. Such outcomes will also allow the establishment, and initial evaluation, of a transversal maturity model to guide the onboarding, uplifting and continuous engagement of scientific communities. In addition and with T1.6, they will work to onboard new communities, providing both technical and scientific support to ensure a positive user experience and facilitate onboarding of new users and communities. As part of the on-boarding process, all new communities will be invited to join the Galaxy Propeller project where they will have the opportunity to share their experience, be mentored and accompanied through their journey.

September 2022- August 2023 summary of achievements:

Organizing outreach and onboarding efforts with WP1 helps to mentor and include new scientific communities into the EuroScienceGateway, which is the focus of this task. We organized community-specific events to interact with local and developing user groups as well as coordinate with European projects Skills4EOSC, EOSC FAIR-EASE, EOSC AquaINFRA, EOSC4Cancer, , EuroHPC, EGI, and ELIXIR to offer Galaxy training and mentorship. These initiatives aimed to assist communities willing to make their infrastructure and workflows available via the EuroScienceGateway and to promote the creation of cross-disciplinary workflows by meeting the demands of various scientific fields.

Those involved in this project are recording their experiences of joining and operating inside the Galaxy ecosystem, therefore spotting shared difficulties and techniques that are used to help others. Compiled into the Galaxy Community Propeller onboarding cookbook, these ideas provided new adopters a useful resource. Furthermore, these efforts helped to shape a maturity model that can direct the onboarding process and help scientific communities to remain involved with Galaxy.

Working closely with Task 1.6, the team also offers technical and scientific support to new user communities, therefore guaranteeing a seamless onboarding process and a good experience. Every new community is invited to actively participate in the Galaxy Propeller working group, where they can share their experiences, get mentoring, and gain advantage from a continuous support network as they connect with the EuroScienceGateway infrastructure.

164 https://apc.u-paris.fr/APC_CS/en/acme-kick-eu-funded-project-built-and-astroparticle-and-astronomy-communities-0

165 https://indico.cern.ch/event/1435605/overview

166 https://galaxyproject.org/events/2023-09-06/

September 2023- August 2025 Achievements:

The European project EOSC4Cancer, finished in May 2025, highlighted in their Policy Brief[167] the provision of demonstrators on the use of real sensitive data using Galaxy in the EOSC4Cancer WP3 "*Federated cancer data analysis through portals across Europe*". During this EuroScienceGateway reporting period, three actions stand out to be of impact to the adoption of Galaxy in the Cancer community case.

First, a prototype was presented to transfer data from cBioPortal to Galaxy via API in the session "Connected Galaxy and cBioPortal for launching analysis in Galaxy" in the EOSC4Cancer General Assembly Meeting (2-3 October, 2024). This experience was documented in the EOSC4Cancer deliverable: Chèneby, J., & Hovig, E. (2024). **D3.2 - Integrative analysis of multiple data types within cancer portals**. The report details advancements in integrated cancer data analysis, allowing for the handling of diverse data types within a portable format suitable for federated analysis across different sites. Key developments include the connection between cBioPortal and Galaxy (and vice versa), enabling users to trigger analyses in predefined workflows in Galaxy, which automatically deliver the results of those analyses back to cBioPortal.

Aligned with EOSC4Cancer WP3 mission of accessing and analysing sensitive data, the Biohackathon Europe 2024 project "**Enabling Secure Data Access from Galaxy to (F)EGA**" was introduced to establish a mechanism that allows secure access to the data hosted in EGA, either in its central instance/s or any of its federated nodes, from a Galaxy instance. This involved the development of a functional mechanism that would showcase the processing and accessing of a Crypt4GH dataset in Galaxy based on the minimal prototype previously developed within the scope of the ELIXIR implementation study 2021-2023: "Strengthen Data Management in Galaxy". This strategy is geared towards maintaining data encryption within the Galaxy server at all times, with decryption occurring solely during analysis within the secure compute backend. This process yields the encrypted data result output exclusively accessible to the end-user.

Finally, a second EOSC4Cancer demonstrator has been published in the framework of clinical image analysis combining the processing capabilities of Galaxy with the visualisation capabilities in cBioPortal: Navest, R., Chèneby, J., Achterberg, H., Gammaraccio, F., & Longo, D. (2025). **D2.2 - Analytical methods for data extraction, processing and sharing using biomedical images**. Similarly to the first report, in this demonstrator EOSC4Cancer effectively combines radiology images with genomic and clinical data within cBioPortal in addition to enabling image analysis through a Galaxy workflow using example image analysis. It was showcased that the demonstrated setup is flexible and the image processing could be replaced by any other image analysis workflow compatible with Galaxy.

Aquainfra is another community which expressed interest in adopting Galaxy. The overall objective of the AquaINFRA project is to develop a virtual environment equipped with FAIR multi-disciplinary data and services to support marine and freshwater scientists and stakeholders restoring healthy oceans, seas, coastal and inland waters. The AquaINFRA virtual environment - Aqua's Galaxy - will enable the target stakeholders to store, share, access, analyse and process research data and other research digital objects from their own discipline, across research infrastructures, disciplines and national borders leveraging on EOSC and the other existing operational dataspaces (e.g., EMODnet, Copernicus Marine Service, Digital Twins, etc.).

AquaInfra is developing the processing platform, based on Galaxy and Pygeoapi (a Python platform to run OGC-standard compliant processing services accessible via

[167] https://doi.org/10.5281/zenodo.16572771

REST API). AquaINFRA will process data from diverse sources like pollutant dispersion from rivers to coastal waters in Catalunya, satellite-imagery analysis for the North Sea and the Elbe, nutrients in the Baltic Sea, small river in-situ analysis in Finland, etc. They are looking for HPC resources to run dispersion modeling tasks.

The choice of Galaxy[168] as the platform to create their Virtual Research Environment (VRE) has been determined by the availability of different tools, simple procedures for data import and export, easy, workflow creation, and the fact that results and workflows can be deposited to repositories like Zenodo and get a Digital Object Identifier (DOI) for them. To limit the tools available to a subset that are useful for freshwater and marine researchers, they create a subdomain based on the successful examples of the Climate and Earth system communities.

AquaInfra is represented in ESG WP5 by Markus Konkol (52North, Münster) and Merret Buurman (Institute for Freshwater Ecology), Berlin).

Deliverables
D5.1: Community onboarding cookbook published (M18, Public)
D5.2: Publication of the usage of EuroScienceGateway by multiple communities (M36, Public)
M5.1 Interim report on the activities of the 3 early adopters and assessment of the take up by their respective communities

#§WRK-PLA-WP§# #@IMP-ACT-IA@#

1.3 Impact

Please describe the progress of the project so far towards delivering scientific impact, based on its objectives and towards delivering impact in any of the following fields (if applicable): scientific, economic, societal or industrial production or processes. Report on changes to the expected impacts presented in your DoA (if any) and the effects on the project/need for adaptations.

Where necessary, provide further details of your monitoring and evaluation strategy, including: references to baselines, benchmarks, assumptions used (with justification) as well as calculations performed to quantify the impacts. If necessary, provide this information in a separate deliverable / a dedicated section of a deliverable

General impact

Galaxy for data-intensive research

Pulsar Network

FAIR Digital Objects (FDOs)

Interoperability and sustainability

- Galaxy is the only RIR that is actionable on data in ELIXIR, where users can compute on data

[168] AquaInfra blog is located here https://blog.52north.org/2024/07/25/the-aquainfra-virtual-research-environment-design-rationale-for-using-the-galaxy-platform/

- ELIXIR recommended interoperability resource

Impact on communities

- easier more assessible software, tools, workflows and infrastructure

Monitoring the progress

1.4 Update of the plan for exploitation and dissemination of results (if applicable)

Include in this section any updates to the plan for exploitation and dissemination of results and give details.

#§IMP-ACT-IA§#

[OPTION for projects providing access to research infrastructure:

1.5 Access to research infrastructure

Not applicable.

#§PRO-GRE-PG§# #@FOL-UP-FU@#

2. FOLLOW-UP OF RECOMMENDATIONS AND COMMENTS FROM PREVIOUS REVIEW(S) (IF APPLICABLE)

Please include a table explaining if and how each recommendation from previous reviews and/or Project Officer assessment has been addressed.

#§FOL-UP-FU§# #@IMP-ACT-IA@#

3. EXPLOITATION PRIMARILY IN NON-ASSOCIATED THIRD COUNTRIES (IF APPLICABLE)

Not applicable.

#§IMP-ACT-IA§# #@CON-MET-CM@#

4. OPEN SCIENCE

Describe the Open Science practices related to early and open sharing of research (e.g. through pre-registration, registered reports, pre-prints or crowd-sourcing of solutions to a specific problem).

Describe the concrete measures that ensure the reproducibility of the results obtained during the action i.e., measures to ensure that the *same results* can be obtained by using the *same data and/or methods*, etc.

#§CON-MET-CM§# #@WRK-PLA-WP@#

5. DEVIATIONS FROM ANNEX 1 AND ANNEX 2 (IF APPLICABLE)

Explain the reasons for deviations from the DoA, the consequences and the proposed corrective actions.

5.1 Tasks/objectives

Include explanations for tasks not fully implemented, critical objectives not fully achieved and/or not being on schedule. Explain also the impact on other tasks on the available resources and the planning. Explain also the impact on other tasks and provide and provide details to allow assessing whether the project is on track.

5.2 Use of resources *(n/a for MSCA and Lump Sums)*

Include explanations on deviations of the use of resources between actual and planned use of resources in Annex 1, especially related to person-months per work package.

Include explanations on transfer of costs categories (if applicable).

Include explanations on adjustments to previous financial statements (if applicable).

5.2.1 *Unforeseen subcontracting (if applicable)* *(n/a for MSCA)*

Specify in this section:

- the work (the tasks) performed by a subcontractor which may cover only a limited part of the project
- explanation of the circumstances which caused the need for a subcontract, taking into account the specific characteristics of the project
- the confirmation that the subcontractor has been selected ensuring the best value for money or, if appropriate, the lowest price and avoiding any conflict of interests
- include also the name of subcontractor and amount.

5.2.2 *Unforeseen use of in kind contributions* *(n/a for MSCA and Lump Sums)*

Specify in this section:

- the identity of the third party
- the resources made available by the third party respectively against payment or free of charges
- explanation of the circumstances which caused the need for using these resources for carrying out the work.

#§WRK-PLA-WP§#

6. References

Azab, Abdulrahman, Paul De Geest, Sanjay Kumar Srikakulam, Tomáš Vondra, Mira Kuntz, and Björn Grüning. 2025. “Optimized Meta-Scheduling in Galaxy Using TPV Broker.” Version 1.0. Preprint, Zenodo, February 27. https://doi.org/10.5281/ZENODO.14936846.

Azab, Abdulrahman, Sanjay Kumar Srikakulam, Paul De Geest, et al. 2025. *Publication on the Smart Job Scheduler Implementation*. Version 1. February 27. https://doi.org/10.5281/ZENODO.15827107.

Clark-Casey, Justin, and Stian Soiland-Reyes. 2022. *Making EOSC Research Objects FAIR with RO-Crate: A Common Metadata Overlay for EOSC Repositories*. November 16. https://doi.org/10.5281/ZENODO.7323480.

Erxleben, Anika, Sebastian Schaaf, Flora D’Anna, Bjoern Gruening, and Frederik Coppens. 2023. *EuroScienceGateway Data Management Plan (DMP; D1.1)*. Version 1.0.

October 30. https://doi.org/10.5281/ZENODO.10054551.

Erxleben, Anika, Sebastian Schaaf, Oana Marchis, and Bjoern Gruening. 2023. *EuroscienceGateway Dissemination & Exploitation Plan (D&E; D1.2)*. Version 1.0. October 30. https://doi.org/10.5281/ZENODO.10055554.

Nicotri, Stefano, Marco Antonio Tangaro, Federico Zambelli, et al. 2024. *D3.1 Operations Documentation on the Open Infrastructure Deployment*. August 27. https://zenodo.org/records/15920451.

Pedersen, Maiken, Sanjay Kumar Srikakulam, Paul De Geest, et al. 2024. *Bring Your Own Infrastructure*. Version 1. August 26. https://doi.org/10.5281/ZENODO.15729502.

Gustafsson, O.J.R., Wilkinson, S.R., Bacall, F. *et al.* WorkflowHub: a registry for computational workflows. *Sci Data* 12, 837 (2025). https://doi.org/10.1038/s41597-025-04786-3

Langer, B.E., Amaral, A., Baudement, MO. *et al.* Empowering bioinformatics communities with Nextflow and nf-core. *Genome Biol* 26, 228 (2025). https://doi.org/10.1186/s13059-025-03673-9

Wilkinson, S.R., Aloqalaa, M., Belhajjame, K. *et al.* Applying the FAIR Principles to computational workflows. *Sci Data* 12, 328 (2025). https://doi.org/10.1038/s41597-025-04451-9

Leo S, Crusoe MR, Rodríguez-Navas L, Sirvent R, Kanitz A, De Geest P, et al. (2024) Recording provenance of workflow runs with RO-Crate. PLoS ONE 19(9): e0309210. https://doi.org/10.1371/journal.pone.0309210

De Geest P, Coppens F, Soiland-Reyes S, Eguinoa I, Leo S (2022) Enhancing RDM in Galaxy by integrating RO-Crate. Research Ideas and Outcomes 8: e95164. https://doi.org/10.3897/rio.8.e95164

Harrison E, Clarke Z, Balsyte E and Maccallum P. ELIXIR All Hands Dublin 2023 – Setting the stage for the ELIXIR 2024-28 Scientific Programme [version 1; not peer reviewed]. *F1000Research* 2023, **12**(ELIXIR):1126 (document) (https://doi.org/10.7490/f1000research.1119616.1)

Paul De Geest, Exporting Workflow Run RO-Crates from Galaxy (Galaxy Training Materials). (2024) https://training.galaxyproject.org/training-material/topics/fair/tutorials/ro-crate-in-galaxy/tutorial.html Online; accessed Wed Oct 22 2025

Hambley, A., Woolland, O., Chadwick, E., Savchenko, V., Fernández, J. M., & Soiland-Reyes, S. (2025). *Build WorkflowHub Knowledge Graph*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.1967.1

Hambley, A., Chadwick, E., Woolland, O., Soiland-Reyes, S., & Savchenko, V. (2024). WorkflowHub Knowledge Graph (2024-08-22) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13362051

Chadwick, E., Woolland, O., Hambley, A., Savchenko, V., & Soiland-Reyes, S. (2025). WorkflowHub Knowledge Graph (2025-08-29) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.16995374

Beer, P., Szente, M., & Chadwick, E. (2024). rocrate-inveniordm (v2.0.3). [Computer software] Zenodo. https://doi.org/10.5281/zenodo.13366072