

Galaxy Community Conference 2026

Clermont-Ferrand, France

June 23-26, 2026

Abstract Book

Abstract Listing

Talk & Poster

1. SCHNEIDER Daniela, SUDERMAN Keith. *Natural Language Processing in Galaxy: Integrating Stanford CoreNLP and spaCy for the Digital Humanities*
2. ZIEREP Paul, BATUT Bérénice. *FAIRyMAGs: A Modular, FAIR-Compliant Galaxy Workflow Suite for Flexible and Scalable Metagenome-Assembled Genome Reconstruction*
3. LOACH Marisa, RUE-ALBRECHT Kevin. *FAIR Workflows and Training for Single-Cell Analysis in Galaxy and Beyond*
4. PRICE Gareth, GOONASEKERA Nuwan, BROMHEAD Catherine, MATHER Marius, ZHU Amanda, MOK Winnie, WINTER Uwe, LEE Justin, MAKUNIN Igor, HYDE Cameron, WARD Nigel, MANOS Steven, AMARAPATHY Samitha. *If you love something, set it free. If it comes back, it's yours - A Galaxy perspective on outsourcing user identity management*
5. JOSSÉ Marie, SEGUINEAU Pauline, LE BRAS Yvan, DETOC Jérôme, NORVEZ Olivier, GRELLET Sylvain, RIZZO Alessandro, KEUCHKERIAN Samuel, DELAPORTE Pascal, BODÉRE Erwan, SARRAMIA David, GUIMONT Mathieu. *Advancing Earth System Research Through Galaxy, Data Terra, and EOSC Collaboration*
6. LE BRAS Yvan, SEGUINEAU Pauline, GALAXY ECOLOGY Community. *Galaxy Ecology: 2026 update*
7. AFGAN Enis, SUDERMAN Keith, NARVAEZ-BANDERA Isis, GOECKS Jeremy, SCHATZ Michael. *Galaxy in AnVIL: Evolving Infrastructure for Working with Protected Data*
8. AWAN Ahmed, GRÜNING Björn, LOPEZ David, SCHATZ Michael. *Beyond the Tool Panel: New Features for Improved Tool Discovery and Management*
9. THOMAS Laurent, GIRARDOT Charles, SCHOLTALBERS Jelle, MONFORT Matthias, REZA Nayeem. *LabID Meets Galaxy: Continuous Provenance from Samples to Results*
10. BASHEA Chala, GETU Melak, GEBREMICAEL Gebremedhin, ALI Abraham, MARBURG VIRUS OUTBREAK Task Force, TADESE Gemechu, TOLLERA Getachew. *Phylogenomic and Functional Analysis of Ethiopia's First Marburg Virus Outbreak Highlights a Single Spillover Event and Preserved Vaccine Targets*
11. KUCHER Natalie. *Charting the Course: Developing the First Cross-Project Roadmap for the Galaxy Ecosystem*
12. MATUSEVICIUS Povilas. *From legacy pipelines to reusable Galaxy workflows for national bacterial WGS surveillance at Statens Serum Institut*
13. COLLINS Tyler, MICHELLE Savage. *Evaluating AI Agent Systems in Galaxy: A Live Integration Testing Framework*
14. SAVAGE Michelle, GRÜNING Björn, PRICE Gareth, SCHATZ Michael. *The Global Galaxy Registry: Structured Metadata and Dynamic Discovery Across 180+ Instances*

15. YORK Spencer, JOSHI Jayadev, RAUBENOLT Bryan, BLANKENBERG Daniel. *Open Source Resources for the Automated Generation of Galaxy Tools*
16. JACQUES Pierre-Étienne, DAVIS John, GAUTHIER Carol, FORTIN Jérôme, BARRETTE Michel. *UseGalaxy Canada joining the UseGalaxy.* federation*
17. BAKER Dannon, CHILTON John, VAN DEN BEEK Marius. *Building an AI-Native Galaxy: Agent Operations, Agent Framework, and MCP*

Talk only

1. LARIVIERE Delphine. *Updates from the Global Galaxy Training Network (GTN) and Galaxy Training Academy (GTA)*
2. DEMANGE Fanny, REFAHI Yassin, LOUX Valentin, PAËS Gabriel. *Integration of experimental protocols into the Galaxy platform to support FAIR experimental workflows*
3. PETERA Mélanie, WORKFLOW4METABOLOMICS Coreteam. *A journey around Galaxy on the W4M boat: a feedback from someone caught in the middle*
4. BARNABÉ Agnès, FERNANDEZ Emilie, LE FLOCH Erwan, LACROIX Thomas, SCHBATH Sophie, LOUX Valentin. *Leveraging Galaxy's Superpowers to Enhance Fermentation Innovation through the Siduri portal*
5. RAMASAMY Pathmanaban, TICHSHENKO Natalia, DE GEEST Paul, DEPOORTERE Boris, ANDRADE BUONO Rafael, MARTENS Lennart, VRANKEN Wim. *Launching the Scop3P Toolkit Starship into Galaxy: From Earth-Bound Services to Galactic Interactive Tools*
6. CALLAN Danielle, VAN DEN BEEK Marius, BAKER Dannon, ROGERS David, CAIN Scott, SMEDS Patrik, CLAWSON Hiram, CORAOR Nate, BEAVERS Kelsey, HAEUSSLER Maximilian, SCHATZ Michael, POND Sergei, NEKRUTENKO Anton. *Starting from Data: Connecting External Resources and Analysis in Galaxy*
7. GUERLER Aysam, CHILTON John, AWAN Ahmed, BAKER Dannon, VAN DEN BEEK Marius, HEIDARI Alireza, LOPEZ David, SAVAGE Michelle, GRUENING Bjoern, NEKRUTENKO Anton, SCHATZ Michael C.. *The Galaxy History Graph: From State to Structure*
8. DELAPORTE Pascal, KEUCHKERIAN Samuel. *Galaxy at BRGM: From Strategic Integration to Operational Challenges in a National Research Infrastructure*
9. VAN DEN BEEK Marius, BROMHEAD Catherine, BERNT Matthias, CHILTON John, GRÜNING Björn. *Evolving TPV: recent advances in resource-aware scheduling for Galaxy*
10. VAN DEN BEEK Marius, DELISLE Lucille, BERNT Matthias, MAIER Wolfgang, LARIVIERE Delphine, IWC Contributors. *Workflows in the Age of LLMs: Building a Galaxy Workflow Community*
11. VAN DEN BEEK Marius, BAKER Dannon, CHILTON John. *User-Defined Tools: From LLM-Generated Analysis to Reproducible, Validatable Workflows*

12. CHILTON John, VAN DEN BEEK Marius, AWAN Ahmed, NEKRUTENKO Anton. *Reproducibility in the Age of Agents*
13. HOJAT ANSARI Mina. *Galaxy-based workflows for genome-resolved and multi-kingdom microbiome analysis: application to the nasal microbiome in Alzheimer's disease*
14. WATSON Greg. *Leveraging Galaxy Across Multiple Scientific Domains*
15. KOSTRYKIN Leonid, WOLLMANN Thomas, GAO Qi, ROHR Karl. *Galaxy Image Analysis: A Decade of Development for Cell Microscopy Image Analysis and Beyond*
16. CORAOR Nate. *Modern Analysis Scales, And So Too Must Galaxy*
17. DURNÍK Robin, HECHT Helge, BABICA Pavel, SOVADINOVÁ Iva, BAJARD Lola. *AOP-toolkit: Galaxy-enabled LLM pipelines for large-scale toxicological literature mining with secure credential management*
18. BAKER Dannon, GRÜNING Björn. *Rewriting the Galaxy Hub -- Migrating 4,300 Pages from Gridsome to Astro*
19. BAKER Dannon, VAN DEN BEEK Marius, CALLAN Danielle, ROGERS Dave, NEKRUTENKO Anton. *From Research Question to Running Workflow: AI-Guided Analysis in BRC Analytics*

Posters

1. SALEEM Hiba, LATEEF Mehreen, ZEHRA Sitwat, ABBAS Tanveer. *Next Generation Sequencing Based Insights Of Biofilm Producing Bacteria: A Step Forward To The Anti-corrosion Strategies For Marine Industry*
2. DUVAUD Severine, GAUTSCHY Rita, NUSSBAUM Johannes. *Introducing Swiss Galaxy for transdisciplinary, reproducible, and open science in Switzerland*
3. ROSIC Nedeljka. *Expanding Marine Natural Product Diversity Through Microbial Genome Mining*
4. BOGAERTS Bert, VAN BRAEKEL Julien, VAN UFFELEN Alexander, D'AES Jolien, GODFROID Maxime, DELCOURT Thomas, KELCHTERMANS Michael, MILIS Kato, GOEDERS Nathalie, DE KEERSMAECKER Sigrid C. J., ROOSENS Nancy H. C., WINAND Raf, VANNESTE Kevin. *Galaxy @ Sciensano – 2026 update*
5. KHALED Ramiz, HERISSON Joan. *Synthetic biology's progress in Galaxy*
6. TOLCHKOV Vladimir, NENOVA Rumyana, TOMOVA Iskra, IVANOV Ivan, DONCHEV Deyan, STOIKOV Ivan, HODZHEV Yordan, DOBRINOVA Yordanka, POCHILIEV Iliyan, MONEV Hristijan, TSAFAROVA Borislava. *Phylogenetic Analysis of Francisella tularensis Outbreak Strains in Bulgaria Using a Reproducible Galaxy-Based Workflow*
7. SRIKAKULAM Sanjay Kumar, FRENTRUP Martinique, LÜBKE Nils-Christian, JOHANNA Nelkner, SCHNITZER Helena, WIBBERG Daniel, MAUS Irena, SCZYRBA Alexander, KOHLBACHER Oliver, DAMMANN-KALINOWSKI Tanja, HOFFMANN Nils. *de.NBI – A National Infrastructure for Data-Driven Life Sciences*

8. CORREARD Solenne, BATUT Bérénice, ELIXIR Biodiversity Community, WATERHOUSE Robert M., BRETAUDEAU Anthony. *The Biodiversity Genomics Galaxy Lab: A Collaborative Hub with curated tools, workflows and training*
9. BATUT Bérénice, NASR Engy, PECHLIVANIS Nikos, STREPIS Nikolaos, ZIEREP Paul. *Microbiology Galaxy Lab: The first community-driven gateway for reproducible and FAIR analysis of microbiological data*
10. BARREAU Arthur, LE BRAS Yvan, LE BRIS Nadine, SEGUINEAU Pauline, GALAXY ECOLOGY Community. *From SAM Segmentation to YOLO Detection: Enabling Semi-Automatic Marine Biodiversity Annotation in Galaxy for Citizen Science Application*
11. RUE-ALBRECHT Kevin, SIMS David, LOACH Marisa, DOYLE Maria, COOLEY Nicholas, RIOUALEN Claire. *BioFAIR Pathfinder: Connecting Bioconductor, Galaxy, and nf-core: FAIR, AI-ready workflows and training for single-cell analysis*
12. LE CRAS Jean, LE BRAS Yvan, SEGUINEAU Pauline, GALAXY ECOLOGY Community. *Facilitating the creation of galaxy tools*
13. PAYET Kevin, DURAND Morgane, GALLUT Cyril, LE BRAS Yvan, SEGUINEAU Pauline, GALAXY ECOLOGY Community. *Updating a ten years old metabarcoding Galaxy toolsuite for Galaxy Ecology and applying to Antarctic conservation needs*
14. MAYI Emilie-Jeanne, LE BRAS Yvan, ELEAUME Marc. *Reusing 15 years old scientific work to guide Galaxy Ecology tools creation for marine spatial planning*
15. VISAN Vlad, BOURDON Vincent, CAILLETAUD Rémi, ROCH Françoise, ROGEZ Yves, THOLLARD Franck. *Galaxy customisation at OSUG: flexible HPC integration & easier workflow migration*
16. GRIFFIN Tim, DO Katherine, MEHTA Subina, WAGNER Reid, JAGTAP Pratik. *Galaxy as an enabling tool for uncovering the dark immunopeptidome through deep learning extensions*
17. GRELET Sylvain, DELAPORTE Pascal, KEUCHKERIAN Samuel. *When Galaxy Meets FAIR Standards: Building Interoperable Virtual Research Environments Across Disciplines*
18. AWAN Ahmed, VAN DEN BEEK Marius, NEKRUTENKO Anton, SCHATZ Michael. *The Ever Evolving Galaxy Workflow Graph: Editor and Invocation View Polish*
19. VAN HEUSDEN Peter, BUHLE Ntozini. *Using a Galaxy Workflow in an ISO-Accredited lab*
20. RANDAZZO Walter, ALCALINE BELLIDO Beatriz, GARCÍA GUTIÉRREZ Enriqueta, BAKER David, NARBAD Arjan, RODRIGO Dolores. *User-friendly genome assembly and annotation workflow on Galaxy platform to track foodborne pathogens in novel seaweed-based food*
21. SUDERMAN Keith, CHILTON John, AFGAN Enis, NEKRUTENKO Anton, SCHATZ Michael. *Dispatching Galaxy Jobs to Google Batch: Direct API vs Pulsar AMQP Sidecar*
22. SEACORD Andrew. *Towards Democratisation of Whole Genome Alignments*
23. LÓPEZ-RODRÍGUEZ Víctor Rubén de Jesús. *Leveraging galaxy tools for clinical variant processing and interpretation in inherited retinal diseases*

24. GARCIA Priscilla, VAN LAAR Tricia. *Genome Analysis Suggests Reclassification of a Putative Clostridium botulinum Strain*
25. KOIRALA Shubham, ZHANG Xiaodan, GURAGAIN Aastha, JOSEPH Dimitri, PALMER Zolt, SHANKAR Rama, LESHCHINER Dmitry, CHEN Bin. *Pan-Disease Transcriptomic Signatures for Translational Target Discovery*
26. GOLPARIAN Daniel, JAKOBSSON Jenny, LEONOR Sánchez-Buso, UNEMO Magnus. *Implementing SensiTyper in Galaxy for personalized gonorrhoea treatment*
27. ATTIQUE Zarlish, ANSARI Adnan Ahmed. *Multiomic Integration and Deep Learning-Driven Network Modeling for Reproducible Cancer Biomarker Discovery*
28. BERULDTSEN Finn, ANTUNES Dinler. *When and Where Proteins Move: RMSX and Flipbook for Molecular Dynamics*
29. NIKITA Nikita, DHIR Sunny. *A Reproducible Galaxy Workflow for Genome Assembly and SNP Analysis of Hibiscus Chlorotic Ringspot Virus (HRS2 Isolate)*
30. LOPEZ TABERNERO David, HEIDARI Alireza. *Reimagining Uploads in Galaxy*
31. TRIPATHI Prabhat. *ABCP_finder: Transformer Embedding-Based Prediction of Anti-Breast Cancer Peptides*
32. MAHMOUD Alexandru. *Administering a Galaxy Instance Tailored for Interactive Workshops*
33. NAGHSH NILCHI Amirhossein, JUMAH Khaled, VIDEM Pavankumar, SPATIAL2GALAXY Consortium, GRUENING Bjoern. *Unlocking Spatial Biology: End-to-End FAIR Workflows for Spatial Data Analysis in Galaxy*
34. KUMAR Anup, SALIKO Aerda, RIETHMUELLER Luca, KADKHODAEI Arash, NILCHI Amirhossein Naghsh, CHIANG Diana, GRÜNING Björn. *AI ecosystem for life science data analysis in Galaxy*
35. VULU Fabien, BOBANGA Thierry, MAMPUYA Pitshou. *Genetic structure of Aedes albopictus in the Democratic Republic of the Congo*
36. MAREUIL Fabien, PLANEL Rémi. *Galaxy@Pasteur: A Centralized Execution Engine for Bioinformatics Services*
37. POLUNINA Polina. *Making a non-filesystem scientific resource browsable in Galaxy: the MaveDB example*
38. LEFEUVRE Hugo, BATUT Bérénice, DIEUAIDE Amanda, SIGURET Cléa, MAREUIL Fabien, LAO Julie, TACKX Raphaël, VELO-SUAREZ Lourdes, HERY-ARNAUD Geneviève, PONS Nicolas, MEDIGUE Claudine, OLIVIER LECLERCQ Sébastien, RUPPÉ Étienne, HAYER Juliette, GLASER Philippe, BIHOUEE Audrey, CHAFFRON Samuel. *Developing Metagenomic Workflows using Galaxy for the National Antibiotic Resistance Surveillance Platform (ABRomics-analysis)*
39. VIDEM Pavankumar, SPOC. *SPOC: Galaxy Single-cell & sPatial Omics Community of practice*

40. HILTEMANN Saskia, CLARK James W., KENRICK Paul, GRÜNING Björn A., DONOGHUE Philip, VRIES Jan de, RENSING Stefan A.. *A Galaxy Workflow for Phylogenetic Analysis and Gene–Species Tree Reconciliation in Land Plant Evolution Research*
41. KA Abdou Lahat, RUIN Quentin, CLADIERE Mathieu, WORKFLOW4METABOLOMICS Coreteam, PETERA Mélanie. *First-time Galaxy tool integration: feedback about challenges faced, from two trainees with different academic backgrounds*
42. UEDA Joe, TEYSSET Laure, CARRÉ Clément, NAOUAR Naïra. *Novel Integrated Galaxy Workflows for comprehensive sncRNA Analysis*
43. HOTZ Hans-Rudolf, LOPEZ-DELISLE Lucille, GOUÉ Nadia, BERNT Matthias, ČECH Martin, DAVIS John, VISAN Vlad. *Small Scale Admins*
44. CORREARD Solenne, BRETAUDEAU Anthony, ZIEREP Paul, BATUT Bérénice. *Navigating Galaxy's Growth: How CoDex Lower Barriers to Access*
45. OLHA Jaroslav, VALKOVSKÝ Viktor, JURÁČEK Jaroslav, KŘENEK Aleš. *Repository-Coupled scRNA-seq Analysis with Galaxy*
46. LOPEZ-DELISLE Lucille, HOTZ Hans-Rudolf, THANKAM SREEDHARAN Vipin. *The Galaxy community in Switzerland*
47. MASSEI Riccardo, BERNT Matthias, NICOLAY Elena Katharina. *Processing zebrafish (Danio rerio) video and image data using Galaxy: A web based solution for toxicology and ecotoxicology research*
48. PRAGASSAM Anthony, CHAUSSEPIED Thomas, BRETAUDEAU Anthony, LE CORGUILLÉ Gildas, FAULON Jean-Loup, LOUX Valentin. *Galaxy-BioProd: An Integrated FAIR Platform for Synthetic Biology, Biotechnology and Life Cycle Analysis*
49. PLANEL Rémi. *GaaS: Galaxy as a Service*
50. LEROI Laura, JOSSÉ Marie, CAER Gwenaél, DETOC Jérôme. *Towards an ODATIS Galaxy Environment for Reproducible Cross-Domain Marine Data Analysis*
51. CECH Martin, DEMKO Martin. *Three Galaxies, One Grid, One Foundation: Building National Bioinformatics Infrastructure in Czechia*
52. SORANZO Nicola, LOACH Marisa. *Launching the Galaxy UK Special Interest Group: National ecosystem coordination supported by the BioFAIR Fellowship*
53. GUNDERSEN Sveinung, CHAVERO DÍEZ María, DE GEEST Paul, DE LA TORRE PERNAS Sabela, MAIER Wolfgang, REPCHEVSKY Dmitry, MOKHTAR FARAG Yehia, LÓPEZ TABERNERO David, NIEDERLÖHNER Ralph, VÁZQUEZ Pável, AASLAND Rein, COPPENS Frederik, GRÜNING Björn, HOVIG Eivind, RAMBLA Jordi, CAPELLA-GUTIERREZ Salvador. *SemSec: Using Crypt4GH Recryption to Secure and Orchestrate Analysis of Encrypted Datasets from EGA/FEGA in Galaxy*
54. BHATI Umesh, SHANKAR Ravi. *ProMaya: a hierarchical universal Deep Learning framework for accurate and interpretable Protein-Protein interaction identification*

Demos

1. MINKOVA Mirela. *Reproducible and Secure Galaxy environments for the EOSC community*
2. GUNDERSEN Sveinung, THE SEMSEC TEAM. *SemSec: Using Crypt4GH Recryption to Secure and Orchestrate Analysis of Encrypted Datasets from EGA/FEGA in Galaxy*

Natural Language Processing in Galaxy: Integrating Stanford CoreNLP and spaCy for the Digital Humanities

SCHNEIDER Daniela, SUDERMAN Keith

Talk & Poster

Galaxy has a long and successful history in bioinformatics and the natural sciences. Its domain-agnostic platform design, however, extends well beyond these fields. One of the latest Special Interest Groups (SIGs) has emerged from the Digital Humanities (DH) and Social Sciences, where researchers increasingly require computational tools for analyzing large text corpora but often lack programming expertise. In this talk, we present the integration of two leading Natural Language Processing (NLP) frameworks into Galaxy: Stanford CoreNLP and spaCy.

Stanford CoreNLP is a Java-based NLP toolkit with a unified annotation pipeline; our Galaxy wrapper exposes its full pipeline in eight languages. spaCy is a Python library that delivers fast, accurate processing using pretrained models for more than 25 languages. By supporting multiple languages, both tools help overcome the English-centric bias of many Digital-Humanities applications, enabling a more inclusive, multilingual research environment.

Both tools run in Docker containers for consistent execution across deployment environments and produce output in common NLP standards like JSON, CoNLL, CONLL-U, and plain text. Language models are managed through Galaxy's data manager framework and presented as selectable options in the tool interface. This design means researchers can perform sophisticated NLP tasks—analyzing sentiment in historical documents, extracting named entities from literary corpora, or parsing syntactic structure across multilingual text collections—without writing a single line of code. Introducing those two tools, which normally run in separate environments, allows users to quickly compare which tool is most accurate for their analysis without downloading anything or learning to code.

Moreover, integrating Stanford CoreNLP and spaCy into Galaxy connects those tools to the full Galaxy ecosystem. Researchers can leverage them within fully reproducible, shareable workflows that interoperate with thousands of existing Galaxy tools.

A typical workflow might involve (i) bulk upload of images, (ii) optical character recognition (OCR) to plain text, (iii) token-level processing with spaCy or CoreNLP, (iv) sentiment analysis, (v) extraction of persons, places, and organisations via named entity recognition (NER), and (vi) visualisation or statistical modelling with Galaxy's various visualisation tools.

Researchers could upload archival material, use Tesseract to extract text from the images, and then clean the text with regular expression tools. The NLP tools allow for extracting locations from administrative files and visualising, for example, how the focus of foreign policy shifted. Instead of doing this manually, Galaxy enables “distant reading” to generate insights for hypothesis testing or large-scale information extraction.

Sharing workflows is easy, allowing domain experts to incorporate best practices for users less familiar

with the tools. Each workflow can be executed on high-performance compute clusters, version-controlled, and exported with a single click—fulfilling FAIR data-management principles and facilitating transparent peer review.

By unifying OCR and advanced NLP within a single, reproducible environment, the new Galaxy tools lower technical barriers, accelerate hypothesis testing, and open up vast corpora of previously inaccessible material to quantitative inquiry. This is a foundational step for the DH SIG, and we welcome collaboration with the Galaxy community to expand the range of NLP capabilities available on the platform.

FAIRyMAGs: A Modular, FAIR-Compliant Galaxy Workflow Suite for Flexible and Scalable Metagenome-Assembled Genome Reconstruction

ZIEREP Paul, BATUT Bérénice

Talk & Poster

Advances in whole-genome sequencing (WGS) have enabled large-scale recovery of metagenome-assembled genomes (MAGs), offering unprecedented insights into microbial diversity across environments. Despite these advances, MAG reconstruction remains computationally intensive and methodologically complex, often requiring the integration of multiple specialized tools for quality control, assembly, binning, refinement, and annotation. Existing workflows frequently rely on scripting-based implementations, limiting accessibility, reproducibility, and adaptability for many researchers and community-driven initiatives.

To address these challenges, we present FAIRyMAGs, a FAIR-compliant, modular workflow suite built entirely within the Galaxy platform for generating and analyzing MAGs. FAIRyMAGs is supported by the ELIXIR Biodiversity, food security and pathogens (BFSP) programme and developed in collaboration with four ELIXIR nodes.

FAIRyMAGs encompasses six interconnected workflows covering all key steps of MAG reconstruction, including read preprocessing, host contamination removal, assembly, binning, dereplication, and downstream taxonomic and functional annotation.

By leveraging Galaxy’s graphical interface, federated compute infrastructure, and robust workflow management, FAIRyMAGs enables researchers to perform complex analyses without requiring local installation or programming expertise. Its modular design allows flexible adaptation, iterative optimization, and seamless integration of new community-contributed tools, empowering both novice and advanced users. In particular, this modular approach, rather than a rigid end-to-end workflow, represents a key advantage of the implementation, and we highlight it here as a model for building more flexible, reusable, and community-extensible Galaxy workflows.

Furthermore, integrating a framework into Galaxy to benchmark different bidders and bin refiners using the CAMI infrastructure enables the identification of optimal conditions for specific biomes. This approach moves away from the common “more is better” strategy in many pipelines, reducing unnecessary computational effort and its associated environmental impact in high-throughput analyses.

We validated FAIRyMAGs using the CAMI II plant-associated benchmark dataset, demonstrating performance comparable or superior to established pipelines such as MGnify, nf-core, or MAGNETO. Additionally, application to four real-world microbiome datasets, spanning host-associated and environmental systems, revealed significant variability in MAG recovery, community complexity, and clustering structure, highlighting the importance of adaptable workflows tailored to dataset-specific characteristics.

To further enhance efficiency, FAIRyMAGs incorporates a machine learning-based framework to predict memory requirements for metagenomic assembly, enabling more informed and efficient resource allocation in large-scale analyses. However, current limitations in leveraging workflow outputs for dynamic resource selection within Galaxy's job scheduling (e.g. TPV-based allocations) prevent this approach from reaching its full potential. This highlights an important opportunity for future development and community discussion within Galaxy.

FAIR Workflows and Training for Single-Cell Analysis in Galaxy and Beyond

LOACH Marisa, RUE-ALBRECHT Kevin

Talk & Poster

Single-cell omics enables researchers to investigate life at the scale of the smallest biological unit. However, researchers can be deterred from using this technique by the confusing array of tools and the apparently complex analysis process. As part of BioFAIR's efforts to develop FAIR infrastructure in the UK, we are making single-cell workflows more Findable, Accessible, Interoperable, and Reusable across the Galaxy, Bioconductor, and Nextflow communities.

Connecting the fragmented landscape of single-cell tools across these three platforms will make it easier for users to Find and Access the tools they need. Developing consistent workflows and training materials will make analyses more Interoperable and enable users to move more easily between platforms. Alongside efforts to align single-cell metadata across platforms, a BioFAIR Pathfinder project is using exemplar single-cell workflows to demonstrate how Bioconductor tools can be adapted into Galaxy wrappers and nf-core modules. New workflows will demonstrate the improved interoperability between Galaxy and Bioconductor single-cell tools.

The final component in FAIR is Reusability. Making tools and workflows available isn't enough to ensure that researchers will actually reuse them, particularly in a field such as single-cell that can be intimidating for life scientists with limited computational experience. Annotated Galaxy workflows coupled with hands-on Galaxy Training Network (GTN) tutorials can introduce these users to single-cell analysis. However, taking the next step and applying these workflows to new research questions can still be challenging. A series of online workshops, targeted at students and Early Career Researchers (ECRs) from institutions with limited bioinformatics support, is guiding users through this transition. Feedback from these workshops will be used to improve the GTN tutorials to better enable future users to adapt and reuse Galaxy workflows for their own purposes.

Both BioFAIR single-cell projects will unite in September for a Galaxy-Bioconductor Cofest and Bring Your Own Data event at The Open University. Participants of the single-cell workshops will be supported through their first independent analysis and encouraged to make their own contributions to the GTN. Meanwhile, Galaxy and Bioconductor contributors will collaborate to ensure these users have the tools and workflows they need, whether they use Galaxy, Bioconductor, or move between platforms.

If you love something, set it free. If it comes back, it's yours - A Galaxy perspective on outsourcing user identity management

PRICE Gareth, GOONASEKERA Nuwan, BROMHEAD Catherine, MATHER Marius, ZHU Amanda, MOK Winnie, WINTER Uwe, LEE Justin, MAKUNIN Igor, HYDE Cameron, WARD Nigel, MANOS Steven, AMARAPATHY Samitha

Talk & Poster

Galaxy Australia recently migrated user identity management to BioCommons Access, a federated, single-sign-on identity and access provider for the Australian BioCommons. The goal was to support stronger identity assurance, fine-grained access and quota controls, and smoother movement of data between trusted research services. What sounded like a straightforward SSO migration was, in practice, a deep reworking of how Galaxy Accounts relate to externally managed identities.

The migration was simultaneously actioned by another institutional partner, making the likelihood of account collisions higher, adding to the complexity of the move. Migration touched approximately 50,000 legacy accounts, including duplicates, thousands of which required action related to username or email addresses. Moving this user base into a federated identity model required bulk transformation, manual curation, extensive user communication, and new operational processes for trust, uptime, and support across organisational boundaries.

The work also drove upstream changes in Galaxy's OIDC support. Contributions included configurable extension points for OIDC authentication flows, access-token decoding in the authentication pipeline, synchronisation of provider-managed email and username attributes back into Galaxy. Crucially, the work also helped to unify Galaxy's bifurcated OIDC implementations for Keycloak and Python Social Auth (PSA), providing a backward-compatible, unified implementation of OIDC based solely on PSA, which greatly simplifies Galaxy's codebase and improves maintainability going forward. Together, these changes make Galaxy better suited to services that rely on centralised identity and access management rather than local account ownership alone.

We will present both the technical and procedural lessons from this work: large-scale account reconciliation, modernising legacy identity data, adapting user profiles and login flows, redefining support boundaries when authentication becomes a shared responsibility, and maintaining trust and operational continuity during a disruptive change. After invalidating more than 21 million user sessions, users returned, and Galaxy Australia continued with a stronger foundation for interoperable identity.

This talk will be relevant to Galaxy operators and research infrastructure providers considering SSO or federated identity and access management, especially those facing legacy accounts, imperfect user data, and the practical challenges of integrating Galaxy with externally managed identity systems.

Advancing Earth System Research Through Galaxy, Data Terra, and EOSC Collaboration

JOSSÉ Marie, SEGUINEAU Pauline, LE BRAS Yvan, DETOC Jérôme, NORVEZ Olivier, GRELLET Sylvain, RIZZO Alessandro, KEUCHKERIAN Samuel, DELAPORTE Pascal, BODÉRE Erwan, SARRAMIA David, GUIMONT Mathieu

Talk & Poster

Three years after its first stumbling steps, the Earth System community within Galaxy continues to grow. Initiated within the 3-year project EOSC FAIR-EASE, and building upon the Galaxy-ecology works, the European subdomain Galaxy for Earth System Sciences (GESS) is now evolving with the Data Terra Research Infrastructure. Data Terra is dedicated to Earth observation and environmental data plays a central role in ensuring GESS's sustainability and growth both within the French scientific community and across Europe.

GESS is designed to address the needs of Data Terra's partners, offering a dual instance: a French national platform and a European one. This architecture enables researchers to access tailored resources and services, supporting both local and international research activities. To strengthen this infrastructure, Data Terra collaborates closely with IFB research infrastructure (the French ELIXIR node). GESS now benefits from a dedicated subdomain on usegalaxy.fr, further integrating Galaxy's capabilities into Earth System Sciences.

As an EOSC (European Open Science Cloud) node, Data Terra contributes to a federated network of services and fosters collaborations with other EOSC nodes. GESS enables interdisciplinary research and data-driven innovation. These collaborations are creating new opportunities for scientists to leverage Galaxy's reproducible and scalable analysis framework in Earth System Sciences.

Key use cases illustrate the impact of GESS:

- Dead Trees (in collaboration with NFDI4Earth): Analyzing satellite and field (in situ) data to understand forest health and climate change impacts on ecosystems.
- Hunga Tonga eruption: integrating multi-source Earth observation data to analyse the eruption's global atmospheric and climatic effects.
- Cross-disciplinary Research: Supporting studies that combine environmental, geological, and biological data to address complex challenges like biodiversity loss or resource management. Data Terra's five thematic hubs—AERIS (atmosphere), FORMATER (solid Earth), ODATIS (ocean), PNDB (biodiversity), and THEIA (land surfaces)—provide datasets and use cases, such as studying the interplay between land-use changes, pollution, and ecosystem dynamics.

By bridging national and European efforts, Data Terra and GESS are contributing to a more interconnected and FAIR research ecosystem. This abstract invites the Galaxy community to explore how GESS and Data Terra are shaping the future of Earth System Sciences through collaboration, innovation, and open science.

Galaxy Ecology: 2026 update

LE BRAS Yvan, SEGUINEAU Pauline, GALAXY ECOLOGY Community

Talk & Poster

Since 2018, the Galaxy Ecology initiative has allowed Biodiversity data related research fields to rely on an accessible, reproducible and transparent solution to share analytical processes.

For more than 6 years, Galaxy Ecology initiative has been devoted to integrating into the Galaxy ecosystem a minimal basis of tools, workflows and tutorials to deal with Biodiversity data. In 2025, we finally published our founding article as a guidance framework to apply best practices in ecological data analysis, demonstrating the relevance of the atomization-generalization framework developed. Last year was also an amazing year for Galaxy-E as several projects joined our effort and we accompanied the growth of the Galaxy Earth System Science sister community at both national and European levels. Finally, within a year, important work was initiated on several topics, such as interoperability with the Jupyter Notebook ecosystem or facilitating AI-based scientific methods used by citizen science projects.

We propose here to look back on the richness of the past year and present ongoing Galaxy Ecology activities so we can showcase all the benefits we are gaining thanks to Galaxy Communities and how we are trying to also make others benefit from our work.

Galaxy in AnVIL: Evolving Infrastructure for Working with Protected Data

AFGAN Enis, SUDERMAN Keith, NARVAEZ-BANDERA Isis, GOECKS Jeremy, SCHATZ Michael

Talk & Poster

AnVIL (anvilproject.org) is a national-scale cloud environment for genomics research supporting analysis of large and controlled-access human data, with Galaxy available as one of its end-user analysis environments. At this scale, secure, data-proximal computation is essential: studies such as population-level variant analysis, cross-cohort cancer re-analysis, and consortium-scale RNA-seq can be performed within governed workspaces without redistributing protected data. For the Galaxy community, AnVIL is not simply another public Galaxy server, but a demonstration of how Galaxy can be operated within a platform built for governed analysis of protected datasets. Without access to such a secure Galaxy environment, researchers are often pushed into fragmented workflows that mirror controlled data to local infrastructure, externalize access control from the analysis system, and reduce reproducibility through environment drift and ad hoc execution. Over the past year, Galaxy support in AnVIL has advanced in two main ways: a new lower-cost deployment model and a growing set of AnVIL-tailored workflows for working with data already hosted in the platform.

The new deployment model is centered on a small virtual machine that runs Galaxy itself, while most user jobs are dispatched to Google Cloud Batch for execution. Galaxy on AnVIL has long been provided through per-user launches, but the earlier model relied on a fixed-size GKE cluster that had to run both Galaxy and the analysis workloads. In practice, that meant users had to launch a cluster sized for the largest job they expected to run, which could substantially increase the cost of routine use. The new approach separates the Galaxy service from elastic compute capacity, allowing new instances to start more quickly, lowering cloud costs, and simplifying operations and maintenance while preserving the familiar Galaxy interface and workflow experience. It also fits well with AnVIL's broader model of bringing computation to protected data rather than moving protected data across environments. These advances have transformed Galaxy on AnVIL into a responsible and scalable computing platform that minimizes costs while providing large computing resources when needed.

In parallel, we are developing a set of Galaxy workflows tailored for AnVIL data and usage patterns, focusing on commonly used analytical pipelines: RNA-seq differential expression, pathway enrichment, and machine learning workflows for tabular and image-based data. Unlike general-purpose workflows from the Galaxy's Intergalactic Workflow Commission (IWC), these are designed for seamless integration with AnVIL-hosted datasets. In practice, this enables direct analysis of controlled-access data from AnVIL workspaces without data movement and pre-configured parameters for such

data. These workflows emphasize reproducibility, scalability, and ease of use, serving as practical templates for secure, data-proximal genomic analysis.

In this presentation, we will describe these recent changes from the Galaxy project's standpoint and discuss what they show about operating Galaxy within a modern protected-data platform. More broadly, we aim to introduce AnVIL as both a working environment for controlled data analysis and a useful model for supporting secure, scalable analysis in settings where data governance requirements are as important as raw compute capacity.

Beyond the Tool Panel: New Features for Improved Tool Discovery and Management

AWAN Ahmed, GRÜNING Björn, LOPEZ David, SCHATZ Michael

Talk & Poster

Finding the right tool for your analysis or workflow is at the heart of Galaxy's user experience. Within the Galaxy interface, the Tool Panel has always been the main and central hub for users to find tools that exist in the server, since the platform's inception. Several improvements have been made to that panel in that time-frame which have made tool discovery within that minimal interface more efficient. However, that minimal panel's instant search is not always sufficient to discover new tools without prior tool knowledge, while it is also not personalized enough to quickly access frequently used tools. In this talk, we will highlight two of Galaxy's latest features that have been designed to solve these problems and re-imagine how users discover and access tools in Galaxy.

The first of these is a rich Tool Discovery View, which provides an extensive interface for users to discover new tools in Galaxy. This view is driven by a search that is more extensive because it leverages the back-end which accesses more metadata about tools than the instant (front-end) panel search. This view also organizes and categorizes tools based on EDAM ontologies (<https://edamontology.org/>). EDAM is a comprehensive ontology of bioinformatics operations, data types, topics and formats, and using this information, each tool is displayed in modern cards with detailed information about the tool, including its description, categories, and various tags.

While we've worked on making the discovery of new tools easier, we also wanted to make it easier for users to access tools they already know and use frequently. To that end, we have added the ability for users to personalize their tool panel with an all new "My Tools" panel view, which is designed to be a personalized hub for users to quickly access the tools they use most often, without having to navigate through the entire tool panel.

Future plans for this feature intend to leave behind the strict panel-section-tied grouping of tools in the interface, and instead lean on tool tags which will provide a much more flexible way of grouping tools due to the multiplicity of tags that can be assigned to each tool. We also aim to allow the sharing of tools with other users, to help create shareable tool sets, as rated by actual user needs rather than admin-curated categories.

The core mission behind these improvements is to make tool discovery more in-depth, while also bridging the gap between admin-curated organization and user-specific needs. We hope to share these features with the Galaxy community and get feedback on how we can further improve the experience of discovering and accessing tools in Galaxy.

LabID Meets Galaxy: Continuous Provenance from Samples to Results

THOMAS Laurent, GIRARDOT Charles, SCHOLTALBERS Jelle, MONFORT Matthias, REZA Nayeem

Talk & Poster

Scientific laboratories generate diverse datasets across sequencing, imaging, and proteomics platforms. Connecting structured data management to reproducible analysis in Galaxy can involve manual transfers, metadata loss, and incomplete provenance tracking.

Lab Integrated Data (LabID) is an open-source (gitlab.com/lab-integrated-data/) and FAIR data management platform developed at EMBL that goes beyond electronic lab notebooks. It manages samples, assays, and datasets within a unified system, capturing their relationships to build a continuous provenance chain from biological specimens to raw data while maintaining a tight connection to scientists' notebook. LabID natively supports diverse technologies, ranging from sequencing to imaging, and can be leveraged to manage any experimental setup following the ISA model. This enables efficient handling of complex projects in which samples are assayed across multiple modalities.

We present LabID's integration with Galaxy, enabling a seamless loop between experimental data management and computational analysis. Data can be sent from LabID to Galaxy for processing, and analysis results are imported back into LabID, where they are linked to the original samples, assays, and workflows. This closes the gap between experimental and computational steps, ensuring that derived datasets retain full context and traceability and can be efficiently shared.

Workflows and their executions are captured as first-class objects. LabID can import workflows from Galaxy or WorkflowHub, and import Galaxy workflow invocations to record parameters, inputs, and outputs, extending provenance from raw data through computational processing. LabID adopts the RO-Crate standard for workflow exchange: imported workflow crates are validated, versioned, and linked to their source repository, while workflow runs, including execution parameters, input and output datasets, can be exported as Workflow Run RO-Crate packages. This enables researchers to publish reproducible analysis records to WorkflowHub or share them with collaborators through any RO-Crate-compatible platform.

LabID provides an on-premise platform where data flows to Galaxy for computation and returns with provenance intact. By unifying experimental and computational metadata, LabID enables reproducible research aligned with FAIR principles.

Further information and a demo are available at <https://labid-demo.embl.de/>

This development is supported by the OSCARS European project under grant agreement No 101129751.

Phylogenomic and Functional Analysis of Ethiopia's First Marburg Virus Outbreak Highlights a Single Spillover Event and Preserved Vaccine Targets

BASHEA Chala, GETU Melak, GEBREMICAEL Gebremedhin, ALI Abraham, MARBURG VIRUS OUTBREAK Task Force, TADESE Gemechu, TOLLERA Getachew

Talk & Poster

In November 2025, Ethiopia reported its first Marburg virus disease outbreak with a 64% case fatality rate. Genomic epidemiology was applied to characterize transmission dynamics, phylogeny, and mutations relevant to vaccine and monoclonal antibody efficacy. Samples were processed at the Ethiopian Public Health Institute (EPHI). RNA was extracted after ethanol inactivation and screened by PCR for viral hemorrhagic fevers and arboviruses. Marburg virus-positive samples were sequenced using the Illumina Viral Surveillance Panel 2 on the NovaSeq X plus and MiSeq platforms. Analyses were performed on the European Galaxy server, including quality filtering (fastp), host depletion (Bowtie2), reference mapping (BWA-MEM2), variant calling and consensus generation (iVar), and lineage assignment (Nextclade). Phylogenetic and temporal analyses were performed using 103 global genomes in BEAST with a Skygrid model. MVD diagnosis was confirmed by RT-PCR, and positive samples underwent full-genome sequencing. Seven high-quality genomes (genome coverage >70%) from the Ethiopian outbreak were successfully sequenced and analyzed alongside 103 global genomes derived from humans and the natural bat host, *Rousettus aegyptiacus*. The Ethiopian sequences showed >99.9% nucleotide identity, indicating low viral diversity. Maximum-likelihood phylogenetic analysis demonstrated that all sequences clustered within the MARV.A.1 clade. They were closely related to historical isolates from the Netherlands (the Leiden strain) and Uganda. The Ethiopian lineage and its closest human isolate (JN408064) diverged from a common ancestor circulating in 1994, whereas the Ethiopian lineage and its closest bat isolates (JX458853 and JX458858) diverged from a common ancestor circulating around 1989. Genomic and epidemiological evidence supported a single zoonotic spillover followed by localized human-to-human transmission. Seventy lineage-defining substitutions were identified, including 10 non-synonymous mutations primarily in the glycoprotein gene, with strong purifying selection ($dN/dS < 1$). The outbreak strain showed high similarity to the reference strain, with conserved monoclonal antibody binding sites and no escape mutations. Overall, the outbreak resulted from a single spillover event with limited viral diversification, supporting the continued efficacy of existing vaccines and monoclonal antibody therapies.

Charting the Course: Developing the First Cross-Project Roadmap for the Galaxy Ecosystem

KUCHER Natalie

Talk & Poster

The Galaxy Project has long thrived as a distributed, community-driven ecosystem, and its growing global adoption has brought an exciting expansion of internationally funded projects and distributed developer teams. With this growth comes a recognized need for greater visibility into development priorities and strategic directions across the Galaxy community, including users, developers, and PIs. As a first step towards better alignment,, the US Galaxy team has developed a year-long, cross-project roadmap spanning the major US-funded Galaxy projects, including Galaxy main (usegalaxy.org), AnVIL, BRC.Analytics, the CFDE Cloud Workspace, ITCR, and beyond. Here we discuss the process, structure, and outcomes of that effort.

Building a shared roadmap across projects with distinct funding, organization, and user communities requires coordination across the Galaxy Steering Committee, Technical Board, and Working Groups. We gathered input from these developer groups and PIs representing the funded projects, synthesized and translated the aims into a prioritized, publicly visible roadmap hosted on GitHub ([gxy.io/roadmap](https://github.com/gxyio/roadmap)). The roadmap gives visibility to near-term aims, strategic focus areas, and community needs, offering a mechanism to improve alignment of activities across multiple projects. We review the roadmap quarterly, reflecting on activities that are completed or in progress to ensure timely work and can forecast sufficient bandwidth for future work. We received very positive feedback on the first roadmap process and plan to improve it in future iterations by gathering input from the broader community on the roadmap, defining activities more concretely, and shaping items into more uniform scope.

We will reflect on lessons learned: what made cross-project alignment difficult, where consensus emerged, and how the roadmap is shaping coordination among teams. We will also discuss how we could incorporate internationally funded projects so that the full breadth of work happening across the global Galaxy ecosystem is represented. This initiative represents a meaningful step toward more transparent governance for the Galaxy ecosystem.

From legacy pipelines to reusable Galaxy workflows for national bacterial WGS surveillance at Statens Serum Institut

MATUSEVICIUS Povilas

Talk & Poster

Statens Serum Institut (SSI) in Denmark performs national public health surveillance across a wide range of bacterial pathogens. To support this work, we have implemented a bacterial whole-genome sequencing (WGS) surveillance workflow in a local Galaxy instance and are preparing to replace a legacy in-house system with this new platform. The workflow is already in active pilot use on essentially all bacterial sequencing samples processed at SSI, running in parallel with the previous pipeline at a scale of about 20,000 genomes per year.

The workflow combines established Galaxy tools for read preprocessing, taxonomic screening, assembly, typing, plasmid analysis, antimicrobial resistance detection and assembly assessment with SSI specific tools and wrappers needed for routine surveillance use. These include custom work to improve support for our needs in tools such as pMLST, as well as Bifrost Bridge, a flexible solution for parsing and combining outputs from multiple tools into a single surveillance ready summary table. The resulting workflow produces a consolidated report spanning QC, assembly, typing, AMR, plasmid and related results, making it suitable for downstream public health interpretation and outbreak support.

A key motivation for this transition is not that Galaxy makes the work simpler in the short term. In practice, Galaxy adds an additional layer of software, configuration and operational complexity compared with running the same tools directly through command-line pipelines and schedulers. Adopting it requires sustained effort, team commitment and time to learn how to work effectively with a large and evolving platform. For us, however, the long-term value is compelling: Galaxy offers a path toward a more maintainable, modular and shareable surveillance workflow that can better benefit from community developed tools and can itself be contributed back to the wider ecosystem.

This talk will present our experience moving a national bacterial surveillance pipeline into Galaxy, with a focus on lessons learned for public health laboratories considering a similar transition. We will describe the current Illumina-based workflow, ongoing work to extend the approach to Nanopore data and our goal of making the workflow, wrappers and associated tools openly available for reuse by the community.

Evaluating AI Agent Systems in Galaxy: A Live Integration Testing Framework

COLLINS Tyler, MICHELLE Savage

Talk & Poster

As large language model-based agent systems are increasingly integrated into scientific computing platforms, ensuring their reliability and scientific appropriateness requires evaluation methodology that goes beyond unit testing. Galaxy's AI system employs a router agent architecture that classifies user queries across multiple distinct pathways — including tool recommendation, error analysis, pipeline orchestration, and history summarization — delegating each to a specialized sub-agent optimized for that task. This work presents a live integration testing framework, called MCPeva,¹ that has been designed to evaluate that system rigorously and reproducibly against real-world user scenarios.

The framework executes against a live Galaxy server using actual large language model calls, with no mocked responses. It comprises two complementary evaluation strategies. Deterministic routing tests assert agent handoff correctness by inspecting structured JSON responses, verifying that each query class reaches its intended agent. An LLM-as-judge scoring approach then evaluates open-ended workflow responses against domain-specific rubrics, where a secondary model assesses whether the agent's answer meets the criteria a domain expert would expect. User scenario coverage is designed to expand continuously across scientific domains, with new test cases added as agent capabilities grow and additional fields are onboarded into the Galaxy ecosystem.

A key feature of the framework is that it allows for multi-model comparative evaluation. A single command executes the full scenario test suite across multiple configured models enabling systematic benchmarking across public, privately hosted, and institutionally accessible models — including those available through the Texas Advanced Computing Center — facilitating rigorous performance assessment across a broad and evolving computational landscape. By applying a consistent evaluation methodology across model providers and deployment contexts, the framework yields reproducible, quantitative comparisons of routing accuracy and response quality.

Collectively, this framework represents a foundational investment in the responsible development of AI-assisted scientific computing within Galaxy. By combining deterministic correctness checks, domain-informed response scoring, and reproducible multi-model benchmarking, it establishes the infrastructure necessary to evaluate agent behavior systematically as models, prompts, and agent architectures evolve. The test suite, dashboard tooling, and rubric definitions are open source and developed within the Galaxy codebase, with the long-term goal of making agentic AI evaluation a routine and community-driven component of Galaxy's development workflow.

The Global Galaxy Registry: Structured Metadata and Dynamic Discovery Across 180+ Instances

SAVAGE Michelle, GRÜNING Björn, PRICE Gareth, SCHATZ Michael

Talk & Poster

The Galaxy platform is deployed across hundreds of independently operated instances worldwide spanning national research infrastructures (usegalaxy.org, usegalaxy.eu, usegalaxy.org.au), domain-specific servers (e.g., microbiology, proteomics, single cell sequencing), and institutional deployments at research consortia across dozens of countries. Despite this scale, discoverability of instance capabilities remain limited which constrains efficient resource utilization and cross-instance collaboration. This work describes a coordinated set of contributions to the Galaxy Hub and Galaxy Instances initiative to address these challenges.

At the core is a YAML-based instance schema that captures canonical metadata across several structured dimensions: access model (public/private/academic-only), supported Galaxy version, domain specialization, geographic location, tool panel composition, and reference data. To facilitate, instance operators submit metadata via a templated pull-request at [gxy.io/instances-join](https://github.com/galaxy-hub/instances-join) against the galaxy-hub repository, enabling transparent community curation and version-controlled provenance.

On the presentation layer, enhancements were made to the Galaxy Hub to support filtering by location, domain, reference genomes, and tool versions, alongside an interactive OpenLayers-powered global map visualization that can all be viewed at [gxy.io/instances-view](https://github.com/galaxy-hub/instances-view). To date, this system has structured metadata for over 180 public Galaxy instances and established a submission pipeline that has reduced onboarding friction for new deployments. Ongoing work focuses on automated metadata refresh via CI/CD workflows, schema versioning to accommodate heterogeneous deployment configurations, and API endpoints that would allow downstream ecosystem services — including alignment with the Special Interest Groups, the Galaxy Training Network, and various workflow repositories.

Beyond the metadata, we describe an architecture for incorporating dynamic operational signals - including availability status via uptime monitoring endpoints, active Galaxy version, and toolset composition derived from ToolShed identifiers in order to move toward a live representation of the ecosystem state. By treating Galaxy instances as first-class, semantically described infrastructure resources rather than static web listings, this work improves computational resource discovery for researchers, reduces duplication across the broader ecosystem, and improve upon the federated Galaxy infrastructure.

Open Source Resources for the Automated Generation of Galaxy Tools

YORK Spencer, JOSHI Jayadev, RAUBENOLT Bryan, BLANKENBERG Daniel

Talk & Poster

Developing Galaxy tools often requires substantial manual effort to create XML tool definitions and companion wrapper scripts, making tool integration a persistent bottleneck for the growth of the Galaxy ecosystem. This talk will describe an emerging approach to reducing that burden through the use of fine-tuned, open-source large language models for the automated generation of Galaxy tool XML files and wrapper scripts. By training these models on existing wrappers, software interfaces, and Galaxy development patterns, it becomes possible to accelerate routine tool creation while improving consistency and lowering the barrier to contribution.

We will focus on how open-source models can support practical Galaxy development tasks, including drafting tool definitions, mapping parameters, constructing command sections, and generating wrapper logic from software documentation or command-line interfaces. We will also discuss the value of open training resources, reproducible workflows, and community-curated examples for improving model quality and making automated generation more reliable. Rather than replacing developers, these models are intended to assist them by reducing repetitive work and allowing more time for testing, refinement, and integration of advanced functionality. Agent-based systems may further enhance this process by allowing models to iteratively generate, test, and refine Galaxy tool wrappers with reduced manual intervention.

We will also briefly highlight two openly available resources that illustrate how wrapper generation can already be partially automated in practice: one aimed at Python-based command-line tools and another, R2G2, designed for R and Bioconductor software. Together, these projects provide useful examples of automation strategies, as well as publicly available resources that can inform future model training and evaluation.

By combining fine-tuned open-source language models with existing community tools and standards, we can move toward a more scalable and sustainable process for Galaxy tool development. This work points to a future in which open infrastructure, open models, and open software development practices make it faster and easier to bring new computational methods into Galaxy.

UseGalaxy Canada joining the UseGalaxy.* federation

JACQUES Pierre-Étienne, DAVIS John, GAUTHIER Carol, FORTIN Jérôme, BARRETTE Michel

Talk & Poster

We will detail the establishment and operational evolution of UseGalaxy Canada (UseGalaxy.ca), a national initiative designed to provide Canadian researchers with access to a free public Galaxy service. Supported by the Digital Research Alliance of Canada (<https://www.alliancecan.ca>), UseGalaxy Canada leverages Cloud and Advanced Research Computing (ARC) resources tailored to the Canadian research landscape while remaining open to international collaborators. As such, UseGalaxy Canada is fully operational and provides services to users from academic and research institutions worldwide.

The timing of the 2026 Galaxy Community Conference (GCC26) serves as an important milestone, marking the formal integration of UseGalaxy Canada (UseGalaxy.ca) into the global UseGalaxy.* federation. Since its launch in 2024, the initiative has undergone a rigorous "ramp-up" phase, scaling its production infrastructure and expanding its core technical team. Through this period of iterative development, UseGalaxy.ca has attained the operational maturity and institutional expertise required to support the anticipated influx of global users and the high-throughput demands associated with the UseGalaxy.* designation.

In this presentation, we will provide a comprehensive overview of the technical and organizational pillars supporting UseGalaxy Canada. This includes the infrastructure architecture and a breakdown of the large-scale Cloud and ARC resources that power the platform's backend to ensure reliability and scalability. We will also present usage statistics that demonstrate the platform's rapid adoption within the Canadian and international scientific communities. We will also offer insights into the outreach strategies, cross-institutional partnerships, and specialized training events that have been instrumental in building a sustainable user base.

Finally, we will present a roadmap for the future evolution of the platform, specifically regarding development, service improvements, computational resources, and team support.

Building an AI-Native Galaxy: Agent Operations, Agent Framework, and MCP

BAKER Dannon, CHILTON John, VAN DEN BEEK Marius

Talk & Poster

Natural-language interfaces are showing up everywhere in bioinformatics, but most are still thin wrappers around a single chat box or a one-off demo. In Galaxy, we are taking a different approach: building AI support into the platform itself so the same capabilities are available to Galaxy users inside the UI and to external AI clients through open standards.

This talk presents Galaxy's emerging AI architecture, which has two pillars. The first is Agent Operations, a new layer built directly into Galaxy and building on the foundational work of the standalone `galaxy-mcp` (<https://github.com/galaxyproject/galaxy-mcp/>) project. This layer provides a shared set of operations for tools, histories, datasets, workflows, jobs, and IWC integration, and exposes them through a built-in MCP server that any compatible client can connect to remotely. The second is an Agent Framework: a registry of specialized agents with intent-based routing, handoffs between agents, and multi-agent orchestration. Both pillars share the same design principle -- they route through Galaxy's existing service layer, inheriting the same authorization, validation, and operational constraints as any other Galaxy client.

ChatGXY, Galaxy's in-application chat surface, ties the two pillars together. It is context-aware: it knows where in the application a user is and routes accordingly. On a tool execution page, it understands which tool is being configured. On a failed job, it can pull stderr and stdout for diagnosis. In other contexts, it can route to specialized assistance such as Galaxy Training Network recommendations, history-aware help, and many other tasks built on the same framework. On the external side, any MCP-compatible client -- Claude Desktop, Cursor, or a custom research agent -- can connect to a Galaxy server and access the same operations: searching tools, invoking workflows, monitoring jobs, and inspecting histories.

This architecture is also a response to a practical problem that many projects are about to face: once AI features start to proliferate, it becomes easy to build several overlapping systems that all need separate maintenance, separate security review, and separate product thinking. By treating AI as shared platform infrastructure rather than as a collection of isolated assistants, Galaxy can add new agents and new clients without rewriting the same capabilities over and over.

We will demonstrate this shared foundation from both sides -- inside Galaxy through ChatGXY and outside Galaxy through an MCP client -- and discuss what we have learned about building AI

infrastructure into a 20-year-old platform without creating parallel systems that will be hard to maintain.

Updates from the Global Galaxy Training Network (GTN) and Galaxy Training Academy (GTA)

LARIVIERE Delphine

Talk only

Here we will present updates on the Galaxy Training Network (GTN), which has developed the Galaxy Training Platform (<https://training.galaxyproject.org>), an open-access, community-driven framework for the collection of FAIR (Findable, Accessible, Interoperable, Reusable) training materials for data analysis utilizing (but not limited to) the user-friendly Galaxy framework as its primary data analysis platform.

Since its inception, this training platform has thrived, with the number of tutorials and contributors growing rapidly, and the range of topics extending beyond life sciences. By providing quality, easily adoptable, openly licensed training materials, the GTN has enabled individuals and educators alike to develop their data analysis skills with ease. By integrating the GTN directly within Galaxy itself, we've managed to bring training directly to students.

To support the instructor community, we provide features to facilitate the use of the materials in a classroom or remote teaching setting, simplifying the contribution flow for new materials, and have added a set of train-the-trainer lessons. The GTN framework has proven to be an invaluable resource for individual learners as well as educators, especially when combined with TIaaS (Training Infrastructure as a Service).

In 2021, during the pandemic, we organised the first worldwide online GTN training event (named Smorgasbord) with hundreds of participants from across the world. Due to its success and popularity, we have repeated this event every year since, now rebranded as the Galaxy Training Academy (GTA).

Over the last three years, the GTA has evolved into a structured, week-long global online training event that builds on the GTN's self-paced materials while fostering community engagement around them. Each year, the GTA offers thematic tracks spanning core Galaxy skills and domain-specific topics, including proteomics, assembly, transcriptomics, single-cell analysis, microbiome research, climate data science, machine learning, and more. Participants follow curated learning paths and benefit from discussion channels and instructor support throughout the week.

Beyond supporting learners, the GTA also helps strengthen the Galaxy training community. It provides opportunities for new instructors to gain experience, encourages collaborative lesson development, and promotes cross-disciplinary exchange. As a recurring global event, it feeds directly back into the GTN ecosystem, helping ensure that materials remain current and responsive to community needs. The GTA

also serves as a key entry point into the Galaxy ecosystem: a significant proportion of attendees are first-time Galaxy users, making it an impactful onboarding pathway for the broader Galaxy community. This reach is only possible thanks to the sustained dedication of the Galaxy community, whose volunteer instructors, developers, and advocates collectively make each edition a success

In this talk, we will present the GTN and the many new topics and tutorials added by the community recently, as well as updates from the latest and upcoming editions of the Galaxy Training Academy and how you can get involved.

Integration of experimental protocols into the Galaxy platform to support FAIR experimental workflows

DEMANGE Fanny, REFAHI Yassin, LOUX Valentin, PAËS Gabriel

Talk only

The sharing of experimental protocols remains challenging, limiting their accessibility and reuse, due to the variability of experimental practices and the lack of standardized protocol representation. To address this issue, it is critical to apply FAIR principles (Findable, Accessible, Interoperable, and Reusable) to improve protocol management. This work focuses on integrating experimental protocols into the open-source Galaxy platform, providing a standardized, executable, and fully traceable environment consistent with FAIR principles.

As this represents a relatively novel approach, it was first essential to establish a robust methodology consistent with the platform's capabilities to ensure effective implementation. At this stage, this approach has been applied to a selected set of key non-confidential laboratory protocols used in biomass transformation (biomass material grinding, chemical composition, and enzymatic hydrolysis), demonstrating its feasibility and practical applicability.

For each protocol, key steps involving critical decisions and parameters for reproducibility were identified and implemented into a dedicated Galaxy tool. Users are invited to provide the required inputs through the tool interface and execute the process, generating a structured record of the entered parameters and decisions for traceability.

The structured record from the first tool is used by a second tool to generate a final report, with optional inclusion of supplementary results such as images or comments. This automated reporting captures all critical decisions and parameters, enabling clear documentation, easy sharing and reuse of experimental protocols.

The method ensures that experimental protocols are traceable, reusable and aligned with FAIR principles. By structuring and recording essential protocols steps, the approach ensures that all critical elements are captured, supporting reproducible experiments.

The strategy could be further revised and optimized to generalize the method, enabling efficient integration of numerous protocols across different domains in Galaxy.

This work was funded by French Research Minister aid managed by the Agence Nationale de la Recherche as part of the France 2030 investment plan through the grant reference ANR-23-PEBB-0008 (GalaxyBioProd project).

A journey around Galaxy on the W4M boat: a feedback from someone caught in the middle

PETERA Mélanie, WORKFLOW4METABOLOMICS Coreteam

Talk only

Once upon a time a group of enthusiastic metabolomics' folks decided to bring themselves together to form Workflow4Metabolomics (W4M), an entity that endeavours to break through the barriers that are obstructing data analysis practices in this field. Metabolomic data analysis is a complex and multistep process, which is constantly evolving with the development of new analytical technologies, mathematical methods, bioinformatics tools and databases. With a similar diversity, the W4M team is composed of various profiles (computer scientist, chemist, biologist...), linked together with Galaxy-based ropes, and it has evolved from an initial tiny group of French engineers to a team with further bounds in Europe.

On this W4M boat, among the eclectic crew that came onboard in 2014, one could find a young engineer in statistics, that was like many other persons in the world of metabolomics: no formal computer science schooling but comfortable with their preferred scripting language, good understanding of the field's data specificities but unaware of other fields' issues, capable of providing precise methodologies to solve data processing struggles but with poor knowledge of community practices to share their work, and being a tool provider and a tool user at the same time.

Fortunately, the Galaxy road that W4M has chosen is one that has a lot to offer. Has the journey been calm and easy? Maybe not always, may you ask the young statistician that sailed with the crew, but Galaxy has strengths that have helped the W4M team to construct, provide and improve an ecosystem around metabolomic data processing that now benefits hundreds of users each year. Galaxy, through its technology, its philosophy, its resources and its community, has enabled the peculiar group that is W4M to provide solutions and support to the metabolomics' community.

Leveraging Galaxy's Superpowers to Enhance Fermentation Innovation through the Siduri portal

BARNABÉ Agnès, FERNANDEZ Emilie, LE FLOCH Erwan, LACROIX Thomas, SCHBATH Sophie, LOUX Valentin

Talk only

Consumers are rethinking their dietary habits and seeking sustainable, durable, safe and healthy foods. Fermentation is a biological process used for years that has proven its potential to naturally enhance bio-preservation, functional capabilities or organoleptic qualities of various food matrices. The French Grand Challenge “Ferments du Futur” (FdF) is a public-private partnership that has been ongoing since 2022. It aims to promote innovation in fermentation using a data-driven strategy to design new fermented products. Each year, FdF funds six projects in various fields such as microbial ecology, sensory and biochemical characterisation, bio-processes design, or host-microorganism interactions. Those projects generate massive and heterogeneous data that can be stored in the user-friendly web portal Siduri available at <https://siduri.migale.inrae.fr/>.

Based on the ontology-driven information system OpenSILEX [1], the web portal Siduri is developed and hosted at the Migale bioinformatics facility. It centralises public data and results from FdF projects to enable data management, interactive visualization and bioinformatics analysis. To provide cutting-edge bioinformatics analysis for fermentation-related data, a catalogue of recommended tools and workflows has been initiated with metadata extracted from bio.tools registry [2] using the EDAM ontology [3]. To stay in line with the FAIR principles, the full traceability of bioinformatics analysis (provenance, parameters, software version...) needs to be tracked when results are uploaded in Siduri.

To tackle this challenge, Siduri takes advantage of Galaxy capabilities. To ensure data privacy, a dedicated instance has been deployed on Migale infrastructure. Tools and workflows identified in the catalogue will be made available on this instance, enabling FdF partners to easily run bioinformatics analysis. Data exchange is performed by an in-house Shiny application that calls both OpenSILEX and Galaxy APIs through their Python client. This application is available from the Siduri portal, allowing users to upload data from Siduri to Galaxy in a new Siduri-tagged history and download datasets from Galaxy to Siduri. Some informations about how the dataset was produced are saved as metadata to ensure traceability, such as the Galaxy workflow ID.

In addition to public tools and workflows, the catalogue can be expanded by applications produced in FdF projects. The Siduri team supports FdF developers in engineering and packaging tools, creating wrappers and designing workflows to share FdF productions in the dedicated Galaxy instance.

The talk will address the challenges associated with Siduri's development and how it leverages Galaxy's ecosystem to perform bioinformatic analyses on fermentation data.

1. Neveu P, Tireau A, Hilgert N, Nègre V, Mineau-Cesari J, Bricchet N, et al. Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. *New Phytol.* 2019;221(1):588–601.

2. Ison J, Ienasescu H, Chmura P, Rydza E, Ménager H, Kalaš M, et al. The bio.tools registry of software tools and data resources for the life sciences. *Genome Biol.* 2019 Aug 12;20(1):164.
3. Ison J, Kalaš M, Jonassen I, Bolser D, Uludag M, McWilliam H, et al. EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics.* 2013 May 15;29(10):1325–32.

Launching the Scop3P Toolkit Starship into Galaxy: From Earth-Bound Services to Galactic Interactive Tools

RAMASAMY Pathmanaban, TICHSHENKO Natalia, DE GEEST Paul, DEPOORTERE Boris, ANDRADE BUONO Rafael, MARTENS Lennart, VRANKEN Wim

Talk only

Proteins perform diverse cellular functions that depend not only on their sequence and structure, but also on dynamic biophysical properties and post-translational modifications (PTMs). Among these, phosphorylation is one of the most prevalent and well-studied PTMs, with important roles in protein regulation, signalling and disease. However, the interpretation of PTM sites in their sequence, structural and biophysical context remains challenging, as relevant analyses are often distributed across separate resources and standalone tools.

To address this bioinformatics challenge, CompOmics and Bio2Byte collaboratively developed the open-sourced Scop3P Toolkit. A set of Jupyter Notebooks that combine phospho-site annotation from Scop3P with protein biophysical feature prediction from B2BTools and interactive visualisation libraries for linear and 3D plots. This toolkit, which is available at <https://github.com/Bio2Byte/Scop3P-notebooks>, supports PTM-centred exploration through integrated 1D, 2D/2.5D and 3D protein views, including linear, topological and residue interaction network representations, together with support for mass spectrometry-derived peptide mapping and mutation-based inspection of protein sequence changes.

This framework has been used in ELIXIR Belgium training activities, like the VIB training “Assess protein dynamics & post-translational modifications through ELIXIR Belgium Node services: B2BTools & Scop3P” (with open access at <https://bio2byte.github.io/training-protein-dynamics-post-translational-modifications>), as well as conference workshops, where it provided a flexible environment for demonstrating multiple protein analysis use cases. Although the positive reception of the tool by participants underscored its relevance and potential value for the scientific community, the Jupyter Notebook application also highlighted the need for a more reusable, accessible and interoperable implementation that could support both training and research in a more professional and stable infrastructure setup.

We present an ongoing collaboration with the UseGalaxy.be team to transform these Jupyter Notebooks into a Galaxy “Interactive Tool”. This effort represents a major milestone given it will be the first Interactive Tool published by the labs guided by the maintainers of the national Galaxy server. The objective is to evaluate how an interactive PTM-focused analysis environment can be adapted to the Galaxy ecosystem and developed into a reusable tool for multiple use case scenarios. The components of the original toolkit were restructured toward an integrated Shiny Application in a Docker container that can be launched directly from Galaxy (image: `bio2byte/scop3p-toolkit:0.1.0`). In parallel, the Scop3P client was integrated into Galaxy as a regular tool, enabling phosphosite-centred data retrieval within Galaxy workflows including modifications, structures, peptides and mutations.

These developments represent an important step toward integrating PTM-centred protein analysis into Galaxy. By combining annotation, prediction and interactive exploration in a single open framework, the Scop3P Toolkit expands the accessibility of ELIXIR Belgium protein resources for the broader proteomics, intrinsically disordered protein (IDP) and 3D bioinformatics communities.

Starting from Data: Connecting External Resources and Analysis in Galaxy

CALLAN Danielle, VAN DEN BEEK Marius, BAKER Dannon, ROGERS David, CAIN Scott, SMEDS Patrik, CLAWSON Hiram, CORAOR Nate, BEAVERS Kelsey, HAEUSSLER Maximilian, SCHATZ Michael, POND Sergei, NEKRUTENKO Anton

Talk only

Galaxy provides a flexible environment for running bioinformatics tools and composing analyses, supporting a range of usage patterns from exploratory work to more structured pipelines. In practice, many users approach problems starting from specific datasets or biological questions, then navigate tools and workflows to get analyses done.

We present a shared framework underlying BRC Analytics (brc-analytics.org) and GenomeArk2 (genomeark2.org) that builds on Galaxy by improving how users discover, configure, and launch analyses in data-driven contexts. BRC Analytics focuses on pathogen genomics, while GenomeArk2 targets vertebrate genome analysis, illustrating how the same approach can be adapted to different communities with distinct analytical needs.

A central idea is to start from data rather than tools—shifting the entry point to analysis without changing the underlying execution model. The framework integrates access to external resources (e.g., UCSC tracks, public read datasets via ENA), along with sequence-based search capabilities (via LexicMap and Logan) that allow users to identify relevant SRA data directly from sequence queries. Selected data can be passed into Galaxy with minimal manual handling, reducing friction around data acquisition and simplifying workflow configuration.

Workflows are drawn from the Intergalactic Workflow Commission (IWC), enabling reuse of community-maintained analyses. We introduce lightweight structure by associating workflows with relevant organisms and use cases, improving discoverability and providing a more guided entry point without restricting flexibility. Users can launch analyses with preconfigured inputs while continuing to rely on the full Galaxy interface for execution and iteration.

We are also exploring LLM-assisted interfaces to support data selection, workflow discovery, and parameter configuration; see the companion talk by Dannon Baker for a detailed discussion of this component.

This approach complements Galaxy's interface, adding a structured layer that better connects data and analysis. We will describe the design decisions, integration patterns, and trade-offs behind this approach, and discuss how it can improve usability, increase workflow reuse, and broaden accessibility for data-oriented users.

The Galaxy History Graph: From State to Structure

GUERLER Aysam, CHILTON John, AWAN Ahmed, BAKER Dannon, VAN DEN BEEK Marius, HEIDARI Alireza, LOPEZ David, SAVAGE Michelle, GRUENING Bjoern, NEKRUTENKO Anton, SCHATZ Michael C.

Talk only

The Galaxy ecosystem continues to evolve, not just in computation and visualization, but in how analyses are represented, queried, and acted upon. We introduce the Galaxy History Graph, a core abstraction that formalizes dataset lineage as a bounded, history-scoped directed graph with explicit nodes and edges for datasets, collections, and tool executions.

Current Galaxy histories expose provenance implicitly as a time-ordered list, where you can trace what produced a given item but not what consumes it, requiring repeated traversal and reconstruction to answer even simple questions about dependencies or workflow structure. This does not scale to complex analyses and cannot support higher-level systems. The History Graph replaces this with a normalized representation where lineage is directly encoded and accessible through indexed relationships within a bounded scope.

This work builds on the transition to structured tool request artifacts and a unified job execution model. The resulting History Graph API provides efficient lineage access through server-enforced node caps, scope-based selection, optional traversal and filtering operations, and a stable identifier scheme. Filtering and dependency resolution become graph operations rather than history iteration, improving performance and simplifying the programming model.

The History Graph establishes a canonical system of record for analysis structure in Galaxy. It enables interactive graph exploration, workflow extraction, and integration with notebooks and reports. It also provides the foundation for AI systems to reason over Galaxy analyses deterministically, replacing ad hoc reconstruction with a stable, machine-readable structural artifact.

In the accompanying demonstration, we present the reference implementation of the History Graph API and its interactive visualization, showing how it unifies provenance access across visualization, automation, and AI-driven analysis in Galaxy.

Galaxy at BRGM: From Strategic Integration to Operational Challenges in a National Research Infrastructure

DELAPORTE Pascal, KEUCHKERIAN Samuel

Talk only

BRGM is deploying an Integrated Digital Platform (Plateforme Numérique Intégrative - PNI) to transform how geological, hydrogeological, and environmental data are accessed, processed, and shared across research projects. Launched in 2025 and fully operational in 2026, the PNI addresses a critical challenge: preventing the proliferation of isolated digital silos while ensuring sustainable interoperability between national and European scientific initiatives.

Strategic Objectives: The PNI aims to centralize, secure, and simplify access to scientific data, tools, and workflows for BRGM researchers and partners. Galaxy has been selected as a core analytical environment within the platform's Data Access Module (MAD), alongside other tools like Cupidon (standardized data catalogs) and Saphir (interactive analysis environments). The platform's modular architecture enables researchers to discover datasets, launch reproducible workflows, and visualize results without managing complex technical infrastructure.

Technical Integration: Galaxy's integration into the PNI leverages its strengths in workflow reproducibility, tool modularity, and community-driven development. The platform connects Galaxy to BRGM's data repositories, high-performance computing resources, and visualization tools, creating a seamless environment from data discovery to publication. The technical architecture emphasizes open standards, hybrid cloud deployment, and compliance with FAIR principles, positioning the PNI as a key node in France's national data infrastructure and the European Open Science Cloud (EOSC).

Cultural and Organizational Challenges: Beyond technical implementation, the PNI represents a profound cultural transformation. Researchers accustomed to desktop-based workflows and local data management must adopt new practices centered on documentation, data sharing, and collaborative platforms. This shift requires sustained change management efforts, including ambassador programs, training sessions, and continuous user support. Resistance to change stems from concerns about learning curves, perceived loss of autonomy, and uncertainty about long-term platform sustainability.

Security and Governance Obstacles: Data security presents significant challenges, particularly for sensitive geological information related to natural resources, hazards, and critical infrastructure. Balancing open science principles with data protection requirements demands robust governance frameworks, access control mechanisms, and compliance with European regulations. The PNI must navigate these constraints while maintaining user-friendly interfaces and efficient workflows.

Collaborative Ecosystem: The PNI is deeply embedded in national and international partnerships. It serves as the digital backbone for several Priority Research Programs and Equipment (PEPR), including IRiMa (Risks), Digital Earth (Subsurface), and One Water. The platform aligns with Data Terra's objectives for Earth system data integration and contributes to EOSC's vision of federated research infrastructures. Through EuroGeoSurveys, BRGM shares best practices and technical

solutions with European geological surveys, positioning Galaxy as a common tool for cross-border geoscience collaboration.

Impact and Perspectives: While concrete user feedback is still emerging, the PNI's deployment marks a strategic shift toward sustainable, interoperable research infrastructures. Galaxy's role extends beyond a technical tool—it embodies a new paradigm for reproducible science, community-driven innovation, and open data practices in geosciences. The platform's evolution will depend on continuous user engagement, technical refinement, and sustained institutional commitment beyond 2030.

This presentation will share lessons learned from deploying Galaxy in a large public research organization, addressing both technical solutions and human factors essential for successful platform adoption.

Evolving TPV: recent advances in resource-aware scheduling for Galaxy

VAN DEN BEEK Marius, BROMHEAD Catherine, BERNT Matthias, CHILTON John, GRÜNING Björn

Talk only

Total Perspective Vortex (TPV) is Galaxy's rule-based scheduling layer for right-sizing job resources and routing jobs to suitable destinations using configurable YAML rules. It allows sites to express policies around tools, users, roles, and destinations, while also reusing a shared community database of default resource requirements and scheduling rules.

This talk will provide an update on the most significant changes to TPV since the v3 series began. A major theme has been stronger configuration validation and maintainability: TPV 3.0 introduced Pydantic-backed configuration support, added type checking to the linter, added linting for unexpected fields, and improved support for multi-file configurations, making large, layered TPV deployments easier to validate and safer to evolve. Later v3 releases continued that work by fixing inheritance handling across multiple config files and adding a merged-config dump, improving confidence for sites that compose shared and local rules together.

A second major theme has been better integration with Galaxy tool metadata and stronger policy controls. Since v3.0, TPV has gained basic support for consuming tool-defined resource requirements and, more recently, support for accepting or rejecting jobs based on `tool_type`. Together, these changes reduce duplication between tool metadata and TPV policy, while giving administrators safer defaults and finer control over how different classes of tools are scheduled.

Finally, TPV has become much easier to debug and explain. Recent additions such as the enhanced dry-run tooling, the new `--explain` mode, and merged configuration dumping make it far easier to inspect why TPV made a scheduling decision and to teach or troubleshoot complex rule sets in production.

Overall, this update will show how the v3 series has pushed TPV beyond a flexible scheduler into a more robust, introspectable, and administrator-friendly tool for operating Galaxy at scale, as well as how the shared db has continued to accrue more sophisticated and scalable rules.

Workflows in the Age of LLMs: Building a Galaxy Workflow Community

VAN DEN BEEK Marius, DELISLE Lucille, BERNT Matthias, MAIER Wolfgang, LARIVIERE Delphine, IWC Contributors

Talk only

In this talk we will describe recent developments and ongoing work within the Intergalactic Workflows Commission (IWC) to support communities in creating high-quality, tested Galaxy workflows that can run anywhere.

The IWC collects and maintains production-ready workflows, lowering the barrier to both contribution and use. We will highlight recent improvements including a redesigned homepage, improved workflow diagrams, and extended validation capabilities. Under the hood, the IWC is moving towards `gxformat2` — a more compact, human-readable, and validatable workflow language — and has expanded testing to include runs against live Galaxy instances with landing requests that include sample test data.

We will also discuss how IWC workflows are being leveraged as integration points in external projects such as `brc-analytics`, broadening the audience for curated workflows beyond the traditional Galaxy user base. Finally, we will explore how large language models are beginning to shape the workflow landscape, from copilot-assisted PR reviews that help maintain workflow quality to early experiments with agent-generated workflows.

Whether you are a workflow developer, a domain scientist looking for ready-to-use analyses, or simply curious about the intersection of reproducible science and AI, come join us and learn how you can contribute.

User-Defined Tools: From LLM-Generated Analysis to Reproducible, Validatable Workflows

VAN DEN BEEK Marius, BAKER Dannon, CHILTON John

Talk only

Last year's introduction of User-Defined Tools provided Galaxy users with a framework to create their own tools using JavaScript expressions and JSON inputs, bypassing the traditional XML tool language and its lengthy review process. By leveraging Pydantic models, JSON Schema, and Monaco editor integration, users gained a streamlined, safe path to extending Galaxy's capabilities on the fly.

This talk expands on that foundation in two directions. First, it addresses the rise of LLM-generated analysis code. Large language models are increasingly used to produce one-off scripts and workflow fragments, but these outputs are typically ephemeral and difficult to reproduce. User-defined tools offer a natural packaging format for LLM-generated code: they provide structured inputs, typed parameters, and containerized execution, turning throwaway snippets into reusable, recomputable components that can be embedded in Galaxy workflows.

Second, the talk presents a graduated path from personal user-defined tools to community-shared best-practice tools. This includes mechanisms for promoting a user-defined tool through stages of peer review, testing, and annotation until it reaches the status of a curated community tool — without requiring the full overhead of the traditional IUC toolshed review process. The goal is to lower the barrier for contributions while preserving quality.

Finally, the talk discusses how the typed, schema-driven nature of user-defined tools enables end-to-end workflow validation independent of the Galaxy runtime. Because tool inputs and outputs are formally described, workflows composed of these tools can be checked for type correctness, parameter compatibility, and structural soundness as static artifacts — making them portable and verifiable across platforms.

Reproducibility in the Age of Agents

CHILTON John, VAN DEN BEEK Marius, AWAN Ahmed, NEKRUTENKO Anton

Talk only

Agent-driven science is scaling faster than our ability to trust it. As agents generate analyses at unprecedented speed, we risk automating the reproducibility crisis itself—producing more results, but with less provenance, less transparency, and more “slop.”

Galaxy needs to meet this moment by confronting a new reproducibility crisis—one that is accelerating with the rise of agents.

The claim we make is simple: the infrastructure built for human reproducibility is exactly what agents need—and **therefore Galaxy’s role must not be to adapt away from its core principles and strengths, but to double down on them.**

We demonstrate this through two new Galaxy capabilities, both designed for humans but that immediately give agents new leverage and incentive to prioritize reproducibility and build workflows that communicate their story and provenance.

History Notebooks are Galaxy-flavored markdown documents attached to histories. They give every analysis a living narrative, with embedded dataset views, job parameter tables, and interactive visualizations that persist and version alongside the data. A researcher can document solo; a human can co-author with the in-app AI assistant; an external agent like Claude Code can drive analysis via the API and document its own work as it goes—building a versioned narrative a human can review after the fact, not just a chat transcript that evaporates. Crucially, when a workflow is extracted from a documented history, the notebook becomes the workflow's built-in report — the narrative travels with the computation, so every future invocation carries the context needed to interpret its outputs. The antidote to slop isn't less automation—it's better infrastructure. There has never been a better time to make the communication of an analysis as reproducible as the analysis itself.

Workflow State Validation brings the static checks of the Galaxy runtime outside the browser and workflow editor — the long-missing piece that makes Format2 (YAML) workflows a first-class citizen. Concise enough for a human to read, structured enough for an agent to compose, and now validated against typed parameter schemas for 10,000+ ToolShed tools. No Galaxy server required. Build in your own editor with autocompletion and inline documentation; validate every parameter name, value, select option, conditional branch, and collection type connection in milliseconds. This is the workflow authoring experience bioinformaticians have loved about text-based systems like Nextflow — but with static guarantees only a centralized typed tool registry can provide. The difference between catching a misspelled output directory and catching an invalid alignment scoring option that would silently produce wrong results. For agents, the leverage compounds: the concise format fits in a context window, structured per-parameter error reports replace trial-and-error execution cycles, and the resulting workflows are statically verified before they ever run.

Together, these features illustrate a broader pattern: Galaxy's reproducibility infrastructure is not overhead for agents to bypass—it is the feedback loop that makes agent-assisted science trustworthy by design.

Galaxy-based workflows for genome-resolved and multi-kingdom microbiome analysis: application to the nasal microbiome in Alzheimer's disease

HOJAT ANSARI Mina

Talk only

Understanding the role of the microbiome in neurodegenerative diseases requires robust, reproducible, and scalable analytical frameworks that integrate taxonomic, functional, and genome-resolved approaches. Here, we present a set of Galaxy-based workflows for comprehensive metagenomic analysis and demonstrate their application to the nasal microbiome in Alzheimer's disease.

Shotgun metagenomic data from 93 individuals, including Alzheimer's disease patients, individuals with mild cognitive impairment, and cognitively healthy controls, were processed using a fully reproducible workflow implemented on the Galaxy platform. The analysis spans the full pipeline from raw sequencing data to downstream interpretation, including quality control and host removal, read-level taxonomic profiling using complementary approaches (MetaPhlAn and Kraken2, harmonized via TAXPASTA), and functional profiling with HUMAnN.

To enable genome-resolved analysis, we implemented an assembly strategy combining individual-sample assembly with diagnosis-level co-assembly using the FAIRyMAGs workflow. This approach enables the recovery of both sample-specific and low-abundance microbial genomes while minimizing assembly artifacts. Across all samples, we recovered 20 near-complete genomes with >98% completeness and <2% contamination.

The workflows further support the detection and analysis of viral components, including dedicated workflows for viral characterization, enabling multi-kingdom microbiome characterization within a unified framework. All analyses are implemented as modular and reusable Galaxy workflows, promoting accessibility, transparency, and reproducibility.

This study highlights the potential of Galaxy-based workflows for scalable microbiome research and demonstrates how such frameworks can support the analysis of complex host-associated microbiomes, including those associated with Alzheimer's disease.

Leveraging Galaxy Across Multiple Scientific Domains

WATSON Greg

Talk only

Oak Ridge National Laboratory (ORNL) delivers scientific discoveries and technical breakthroughs needed to realize solutions in energy and national security and provide economic benefit to the nation. ORNL provides a unique environment for accelerating scientific discovery including the operation of one of the world's most advanced neutron scattering facility, a computational capability that includes two of the largest computer systems ever built, as well as undertaking research many other scientific domains. Over the last five years, ORNL has been developing tools and deploying Galaxy instances in order to better leverage this capability for the analysis of neutron scattering data. More recently, we have started exploring the use of Galaxy in other scientific domains, including fusion energy research and energy-water resiliency.

In this presentation, we will provide an update on our efforts to incorporate Galaxy into multiple scientific domains at ORNL. In particular, we will discuss how our neutron scattering galaxy instance (known as the Neutrons Data Interpretation Platform, or NDIP), combined with our visualization framework (known as the Neutrons Open Visualizations and Analysis framework, or NOVA) is now moving into production use, and our plans to incorporate tools that take advantage of AI model and metadata catalogs. We will also describe new efforts to deploy Galaxy based tools for creating a digital twin of the Material Plasma Exposure eXperiment (MPEX), high-power, steady-state linear plasma device designed to produce the plasma material interaction conditions of the divertor of future magnetic confinement fusion power plants. Finally, we will describe ARMADA, a platform for exploring the resilience of energy infrastructure to ensure an affordable, secure, and resilient supply in the face of rapidly growing demand. To achieve this, we have combined Galaxy with the CKAN data management platform in order to provide a seamless environment for running disparate domain science models for exploring energy resilience.

Galaxy Image Analysis: A Decade of Development for Cell Microscopy Image Analysis and Beyond

KOSTRYKIN Leonid, WOLLMANN Thomas, GAO Qi, ROHR Karl

Talk only

We have developed Galaxy Image Analysis (GIA), a software suite comprising tools and workflows for cell microscopy image analysis that builds upon the Galaxy platform. The development of GIA started 2016 with the goal to establish image analysis in Galaxy, and the software suite was first introduced in a training course in 2016 (Microscopy Image Analysis Course, Heidelberg University). To establish GIA, the infrastructure for provision of tools and corresponding software dependencies in Galaxy (e.g., Bioconda, Grüning et al., Nature Methods 2018) had to be extended by necessary software packages for image analysis.

The GIA software suite has been continuously extended over the years and GIA workflows were used in multiple biological research projects (e.g., Föll et al., GigaScience 2019; Wang et al., Pain 2022). GIA matured as a software suite for cell microscopy image analysis, and the development was accompanied by a growing community who contributed tools and workflows for image analysis to the Galaxy ecosystem (e.g., Kostykin et al., BioHackrXiv 2024; Chiang et al., BioHackrXiv 2026).

A challenge in developing reliable image analysis tools is that the image data often has multiple dimensions of different types (e.g., 3-D spatial dimension, channel dimension, time dimension), which is more difficult to cope with than single-channel static 2-D images. For example, the image metadata needs to be leveraged for correct interpretation of the different image dimensions, and the metadata must also be preserved for intermediate and final results. To streamline the development and sustainability of tools for multi-dimensional image data (including dimensions of different types), we have recently implemented an auxiliary library for GIA (<https://doi.org/10.5281/zenodo.19386399>) that provides a standardized image representation based on the image metadata. In addition, the library seamlessly enables tools and workflows in GIA to directly support TIFF, OME-TIFF, OME-Zarr, and other file formats.

In this talk, we will showcase the challenges of processing image data with multiple dimensions of different types, and the corresponding solutions provided by the GIA auxiliary library. We will also highlight example use cases of image analysis employing GIA. Ten years after GIA pioneered the analysis of cell microscopy images in Galaxy, this will include recent use cases of image analysis beyond cell microscopy. We will also cover the technical requirements that were necessary to be implemented for this extension. In addition, we will report results of an evaluation of GIA.

Modern Analysis Scales, And So Too Must Galaxy

CORAOR Nate

Talk only

The usegalaxy.org public Galaxy service serves thousands of researchers worldwide, distributing computational workloads across a diverse set of HPC and cloud resources located across the United States. Each site brings its own storage systems, unique system configurations, and failure modes. This talk provides an update on the infrastructure and deployment of usegalaxy.org to improve the scalability, reliability, and manageability of this multi-site compute architecture.

A major focus has been the standardization of Pulsar endpoint deployments. Previously, each remote compute site was deployed and configured independently, leading to operational complexity and configuration drift. We have developed a unified deployment framework that brings all HPC Pulsar endpoints under a common configuration and management strategy. This work has also included the deployment of Pulsar Relay, a new communication model in which the Pulsar server initiates connections to Galaxy without the need of an AMQP server, vastly simplifying the deployment and requiring fewer open firewall ports than ever before.

On Jetstream2, the NSF's elastic cloud, scaling via Slurmscale continues to evolve. The Slurmscale deployment process has matured considerably. Recent improvements include support for Interactive Tools on isolated cloud instances, the ability to run instances in multiple projects (bringing us ever closer to "bring-your-own-compute"), and improved utilization of existing resources.

We have also made changes to Galaxy's job lifecycle to improve reliability at scale. When Galaxy dispatches jobs to remote Pulsar endpoints, transient failures - network interruptions, service restarts, or slow metadata operations - can leave jobs in non-terminal states that require manual intervention. New work introduces robustness to make job completion resilient to these failure scenarios, reducing the operational burden of running Galaxy across many distributed compute sites. Additionally, modifications to the process of cleaning scratch space on HPC systems has improved resilience.

Supporting this multi-site architecture requires continuous attention to resource routing and observability. Total Perspective Vortex (TPV) rules are regularly tuned to match tools with appropriate compute resources based on their requirements. CVMFS unprivileged mounting and caching strategies have been developed to smooth out the differences and data access on remote endpoints. And new observability tooling provides better visibility into Galaxy process behavior and system health.

This talk will cover these developments as deployed for usegalaxy.org, with an emphasis on practical lessons for Galaxy administrators managing, or desiring to manage, distributed compute infrastructure for Galaxy.

AOP-toolkit: Galaxy-enabled LLM pipelines for large-scale toxicological literature mining with secure credential management

DURNÍK Robin, HECHT Helge, BABICA Pavel, SOVADINOVÁ Iva, BAJARD Lola

Talk only

Toxicological research relies on extracting mechanistic evidence dispersed across scientific publications, databases, and regulatory documents. While these resources are openly available, manual evaluation remains a major bottleneck. The AOP-toolkit (aoptk) is an MIT-licensed, modular Python framework designed to support automated retrieval, parsing, and analysis of scientific publications in the context of Adverse Outcome Pathway (AOP)[RD1.1][RD1.2][RD1.3] development[RD2.1] - a framework that integrates data from computational models, cell-based assays, and animal studies to predict how a chemical exposure leads to adverse toxicological outcomes. The toolkit offers functionalities for automated retrieval of publications from PubMed and Europe PMC, ingestion of local PDF files, and extraction of structured mechanistic relationships from text, tables, and figures using pluggable LLM backends. Its architecture supports the integration of alternative models and document-processing components, making it suitable for rapid adaptation as AI model capabilities evolve and helping overall data re-use. Additional utilities include chemical name normalization, identifier mapping, and standardized intermediate data formats suitable for downstream pipeline composition.

A set of Galaxy tools wrapping core aoptk capabilities exposes these methods to non-programming users in a reproducible, workflow-driven environment. These tools support literature retrieval, preprocessing, LLM-assisted extraction, and evidence aggregation. To enable secure use of commercial and customizable LLM APIs within Galaxy, the aoptk tools adopt the new Galaxy Credentials system, allowing users to store OpenAI API keys, endpoint parameters, and authentication data required for literature access directly within their Galaxy account. This eliminates the need to embed secrets into histories or workflows and enables site administrators to offer model-specific configurations while maintaining strict credential isolation.

As a validation case, we applied the system to identify chemicals that inhibit gap junction intercellular communication (GJIC), a mechanism of direct cell-cell communication that is discussed as a candidate non genotoxic key event in AOP development for carcinogenesis. When applied to the publication subset used to construct an expert-curated list of 328 chemicals, the automated workflow recovered more than 250 direct matches, demonstrating that LLM-assisted extraction can reproduce substantial portions of expert manual curation at scale. This approach enables systematic evidence aggregation across studies and chemicals, providing mechanistically relevant information directly usable for AOP development and refinement through support for key events and key event relationships. Systematic mapping of chemical effects on AOP relevant key events further enables chemical prioritization for hazard and risk assessment, facilitates identification of reference chemicals for AOP informed NAM development and validation, and supports evidence based identification of assays suitable for key event evaluation based on their differential performance in detecting chemicals with established effects.

Overall, the AOP-toolkit and its Galaxy integration constitute a FAIR, extensible, and secure ecosystem for automated toxicological evidence mining. By combining modern LLM technology with Galaxy's credential management and workflow reproducibility, the system lowers barriers for large-scale literature analysis and enables future extensions toward authenticated, cross-source toxicology knowledge extraction.

Rewriting the Galaxy Hub -- Migrating 4,300 Pages from Gridsome to Astro

BAKER Dannon, GRÜNING Björn

Talk only

The Galaxy Community Hub (galaxyproject.org) is the project's public front door: news, events, the platform directory, regional subsites, and a large body of community documentation all live there. For years it ran on Gridsome, a Vue 2 static site generator that effectively stopped evolving. With Vue 2 end-of-life and the site's GraphQL-driven build becoming increasingly fragile, we undertook a full rewrite to Astro and migrated more than 4,300 pages.

This was not a straightforward framework swap. The legacy site had accumulated years of permissive template patterns, embedded widgets, region-specific content rules, iframe-served bare pages used by Galaxy servers, and URL conventions that outside publications and tools already depended on. Astro gave us a sustainable target, but only if we could preserve those behaviors while moving to a stricter and more maintainable content model.

We will describe the migration strategy that made that possible: a preprocessing pipeline that converts Gridsome-era content conventions into Astro-friendly inputs at build time, a rebuilt search system based on a generated JSON index and client-side filtering, compatibility handling for regional subsites and embedded pages, and a phased cutover plan where old and new builds ran side by side until the Astro version was ready for production.

This is also not the Hub's first migration. The site started as a MoinMoin wiki, moved to Metalsmith, then to Gridsome, and now to Astro. An important theme of the talk is that migrations like this fail when teams treat them as a pure rewrite exercise. The hard part is preserving the undocumented behaviors that users and downstream systems quietly rely on: region-aware content views, iframe-embedded pages consumed by Galaxy servers, legacy query parameters in the platform directory, and a backlog of mixed markdown, raw HTML, and embedded widgets that could not simply be cleaned up by hand.

AI-assisted development also helped make the migration tractable for a small team by handling repetitive content analysis, preprocessing logic, and many of the edge cases spread across thousands of pages.

The result is a Hub with dramatically faster builds, lower dependency risk, and a codebase that new contributors can actually understand and extend. Adding new content is also simpler: most pages remain straightforward markdown, while richer pages use explicit MDX components rather than the older Gridsome-era mix of implicit template behavior and embedded Vue patterns. That makes authoring more predictable for contributors and reduces the amount of framework-specific knowledge

needed to publish new material.

This talk will focus on practical lessons from a large static-site migration in an active open-source project: how to preserve behavior while replacing the stack underneath it, how to handle messy real-world content at scale, and how to de-risk a cutover when the site is too important to take offline.

From Research Question to Running Workflow: AI-Guided Analysis in BRC Analytics

BAKER Dannon, VAN DEN BEEK Marius, CALLAN Danielle, ROGERS Dave, NEKRUTENKO Anton

Talk only

BRC Analytics (brc-analytics.org) is a Galaxy-powered platform for infectious disease and pathogen genomics, built around a curated catalog of organisms, assemblies, and validated workflows. For many bench scientists, however, the hard part is not pressing "run." It is translating a biological question into the exact workflow, reference genome, annotation set, and input configuration that the platform requires. We have been addressing that gap by building an AI-guided interface that turns an imprecise research goal into a validated, executable analysis.

This talk presents the Analysis Assistant, a conversational interface that replaces a rigid step-by-step wizard with guided dialogue. A researcher can begin with a statement such as "I want to analyze my Salmonella whole genome sequencing data," and the system incrementally resolves that request into structured decisions about organism, assembly, workflow, data source, and annotation. The assistant can also search the European Nucleotide Archive (ENA) for relevant public datasets as part of the conversation. As those decisions are made, the assistant tracks the emerging analysis state and hands off to the existing execution flow once the configuration is complete.

The key design choice is that this is not a generic chatbot making scientific guesses. The LLM is used to interpret intent, manage the conversation, and help the user navigate choices, but correctness-critical decisions are grounded in curated platform logic which is covered in depth in our companion talk by Danielle Callan. The assistant matches across taxonomy lineages, identifies workflows compatible with the selected organism and assembly, and deterministically resolves concrete workflow inputs such as reference genomes and annotation files. In other words, language guides the interaction, while validated catalog logic decides what is actually runnable.

We also expose the same catalog and ENA search capabilities through MCP, so the validated knowledge that powers the assistant can be used by external AI clients as well. That makes the project relevant beyond a single interface: it is a model for connecting conversational AI to a domain-specific scientific platform without giving up the structure and determinism that real analysis workflows require.

In this presentation, we will demo the assistant end to end, from an initial research question to a runnable analysis configuration, and discuss lessons from building AI guidance for scientists who need help framing and launching analyses, not just answering questions about them.

Next Generation Sequencing Based Insights Of Biofilm Producing Bacteria: A Step Forward To The Anti-corrosion Strategies For Marine Industry

SALEEM Hiba, LATEEF Mehreen, ZEHRA Sitwat, ABBAS Tanveer

Poster only

Biocorrosion, resulting from the accumulation of microorganisms and materials on submerged marine surfaces, poses significant challenges to the marine industry. This buildup includes biological matter from both prokaryotic and eukaryotic organisms, promoting biofilm formation on structures such as ship hulls and pier pylons. Biofilm development initiates with the adsorption of organic and inorganic particles, followed by pioneering microorganisms attaching, proliferating, and forming primary colonies that mature into a structured biofilm matrix.

The present study aimed to explore the genetic foundations of biofilm formation, a key factor in marine biofouling, using Next Generation Sequencing (NGS) technology. Biofilm-producing bacteria were isolated from organic deposits on submerged ship hulls and screened using the Microtiter Biofilm Assay. Isolates were further characterized phenotypically and genotypically, with genomic DNA quality and quantity verified through gel electrophoresis and Qubit assays.

Phylogenetic variations associated with biofilm formation were analyzed via 16S rRNA sequencing, revealing that among the 10 isolates, two were strong biofilm producers (optical densities of 0.45 and 0.30, with $OD \geq 0.01$ as the biofilm threshold). The 16S rRNA phylogenetic analysis showed 97.97% sequence identity with *Psychrobacter celer*. Whole-genome sequencing of the selected *Psychrobacter* strain on the Illumina platform revealed a 4,128,308 bp circular chromosome with 44.0% G+C content and 4029 coding sequences, with a 92.24% mapping ratio to the reference genome. Genome annotation identified the *cat1* gene, which plays a specific role in biofilm formation and contributes to metal corrosion on ship surfaces.

This study provides molecular insights into the dynamics of marine biofouling and highlights the genetic determinants associated with biofilm formation. The findings advance our understanding of microbial colonization processes on submerged surfaces and offer potential avenues for the development of eco-friendly antifouling strategies to mitigate corrosion and enhance the longevity of marine structures.

Introducing Swiss Galaxy for transdisciplinary, reproducible, and open science in Switzerland

DUVAUD Severine, GAUTSCHY Rita, NUSSBAUM Johannes

Poster only

Many researchers across Switzerland struggle with the steep learning curve of computational tools and high-performance compute (HPC) infrastructures. These challenges create significant barriers to entry. At the same time, reproducibility remains a major concern in research involving computational analyses. Galaxy constitutes a promising solution to address these issues. Our project aims at establishing a Swiss Galaxy instance and integrate key open software tools developed in Switzerland. By doing so, it will democratize access to advanced computational methods, enabling researchers across diverse fields—including life sciences, astrophysics, and the humanities—to perform cutting-edge analyses without requiring specialized computational expertise. At the same time, it will enable impactful “made in Switzerland” tools to be deployed on existing Galaxy instances abroad. By leveraging the multidisciplinary expertise of the project’s partner institutions, this initiative will reinforce Switzerland’s position at the forefront of open, transparent, and reproducible computational research.

Switzerland has long been a leader in developing open databases and tools, particularly in the life sciences, where millions of researchers worldwide rely on Swiss resources. However, integrating these tools into increasingly complex computational pipelines often requires specialized knowledge and poses reproducibility challenges. Furthermore, these tools are often fragmented across disciplines, limiting their potential for transdisciplinary research. This project aims to integrate key Swiss databases and software tools from life sciences, humanities, and astrophysics into a unified, accessible computational ecosystem.

We will port selected high-impact Swiss software tools to Galaxy and deploy a Swiss Galaxy instance on the scalable SWITCH cloud infrastructure, making them available to the Swiss science community. The Swiss instance will meet the specific needs of Swiss researchers, including local datasets, selected tools, and the promotion and testing of Swiss-developed tools. By hosting data and computation within Switzerland, it will also safeguard national capacity. Furthermore, it will incentivize tool developers and users to adopt FAIR and open principles and foster community collaboration and workflow sharing.

By establishing a national Galaxy instance, Switzerland will democratize access to cutting-edge computational tools and workflows, enhancing its leadership in open science, and facilitating transdisciplinary research.

Expanding Marine Natural Product Diversity Through Microbial Genome Mining

ROSIC Nedeljka

Poster only

Marine microorganisms represent a rich and largely untapped source of bioactive compounds with significant potential for biotechnological applications. In their natural environments, most marine species are exposed to diverse biotic and abiotic stressors, including grazing pressure, pathogen exposure, fluctuations in temperature and salinity, and variations in nutrient availability. To survive under these dynamic conditions, marine microorganisms have evolved sophisticated adaptation mechanisms, notably the production of secondary metabolites that enhance their resilience and ecological fitness over evolutionary timescales. Among these metabolites, ultraviolet (UV)-absorbing mycosporine-like amino acids (MAAs) are particularly promising. MAAs are small, water-soluble compounds with absorption maxima in the UVA and UVB range (310–360 nm). They are widely distributed across marine taxa and exhibit diverse bioactive properties, including antioxidant, anti-inflammatory, and photoprotective activities. Despite their considerable potential, the biotechnological exploitation of MAAs has been constrained by low natural yields, variability in production levels, and limited success in heterologous expression systems.

Advances in modern omics technologies have substantially improved our understanding of MAA biosynthesis and its regulatory mechanisms. MAA production typically involves a four-enzyme biosynthetic pathway comprising dehydroquinate synthase, O-methyltransferase, an ATP-grasp enzyme, and a nonribosomal peptide synthetase. However, significant interspecies variation exists in the genes encoding these enzymes, suggesting additional biosynthetic complexity that remains insufficiently characterized.

The application of genome-mining approaches has enabled the identification of genetic diversity within biosynthetic gene clusters (BGCs), facilitating the discovery of novel compounds from previously underexplored sources. Building on existing omics datasets, this study explores the application of advanced genomic technologies to link BGCs with secondary metabolites, including MAAs. Furthermore, it evaluates the potential of innovative genome-mining algorithms to uncover the hidden biosynthetic capacity of marine microorganisms for the development of novel biotechnological products.

Galaxy @ Sciensano – 2026 update

BOGAERTS Bert, VAN BRAEKEL Julien, VAN UFFELEN Alexander, D'AES Jolien, GODFROID Maxime, DELCOURT Thomas, KELCHTERMANS Michael, MILIS Kato, GOEDERS Nathalie, DE KEERSMAECKER Sigrid C. J., ROOSENS Nancy H. C., WINAND Raf, VANNESTE Kevin

Poster only

Galaxy @Sciensano is a free portal for microbial genomics tailored towards public health and clinical diagnostics. It is available to academic and non-commercial users and provides an extensive range of tools dedicated to microbial genomics analysis. Since the launch in 2019, the number of users has grown steadily up until 1,031 at the time of writing.

The platform offers several validated, stand-alone pipelines that can characterize relevant public health pathogens using WGS data, such as *Neisseria meningitidis*, *Mycobacterium tuberculosis*, and *Salmonella enterica*. Moreover, our instance offers many flexible stand-alone tools and databases such as gene detection, (core genome) multi-locus sequence typing, hybrid assembly, etc. These pipeline and tools generate interactive HTML output reports that are easy to interpret for non-experts.

This poster provides an update on the latest developments of our Galaxy instance, namely: (1) full support for input data generated with Oxford Nanopore Technologies (ONT) sequencing, which required updating all pipelines and many stand-alone tools, as well as installation a relevant selection of tools from the Galaxy ToolShed (including the addition and updating of wrappers for tools that were previously unavailable in the Intergalactic Utilities Commission (IUC) tool repository); and (2) an effort to make in-house developed tools open source and accessible to other Galaxy instances. As a proof of concept, several smaller custom tools were contributed to the IUC tool repository, including the phylogenomic tool PACU and the MiST sequence typing tool.

We present a resource that is highly relevant for national reference centers and national reference laboratories in the context of routine pathogen surveillance programs, as well as clinical laboratories, but also research applications. By showcasing it at this conference, we aim to reach a broader audience. Furthermore, by shifting towards open-source development, we aim to promote the integration of custom-developed functionality into other Galaxy instances.

Synthetic biology's progress in Galaxy

KHALED Ramiz, HERISSON Joan

Poster only

Synthetic biology integrates biology and engineering, using advances in genomics and computational tools to design and reprogram living systems. However, the complex data analyses required can be a barrier for biologists without programming expertise. Galaxy, an open web-based bioinformatics platform, addresses this challenge by providing an intuitive interface that enables advanced analyses without coding. Within this ecosystem, Galaxy-SynBioCAD (galaxy-synbiocad.org), a one-stop shop for metabolic pathway design, has been developed. In order to gather tools dedicated to this field, the “Synthetic Biology” category has been created in the Galaxy Tool Shed. This label has since expanded as teams worldwide contribute new tools through the Galaxy infrastructure. At the same time, the synthetic biology landscape continues to grow and in 2019, many biofoundries joined the Global Biofoundries Alliance. In this trend, the Paris Biofoundry was established in 2024, bringing together the DNA Foundry, the Mammalian Cells Foundry, the Cell-Free Biofoundry, and the Scale-Up Biofoundry to coordinate their activities. In the same spirit of Galaxy-SynBioCAD, the Paris Biofoundry has published the tools and workflows used by those foundries (biofoundries.usegalaxy.fr). Some tools are developed in the Biofoundry like in the Cell-Free Foundry, however there are some tools developed by other biofoundries around the world that are useful for Paris Biofoundry like Edinburgh Genome Foundry tools. This dynamic is also reflected in the Galaxy-BioProd project (PEPR B-BEST), where new tools and workflows will be published to address key areas of the B-BEST program (b-best.usegalaxy.fr). The strengthening collaboration between Galaxy and the Synthetic Biology community highlights its growing role as a global platform for accessible and reproducible synthetic biology data analysis.

Mots-Clés: Galaxy-SynBioCAD, Cell-Free Biofoundry, Paris Biofoundry, PEPR B-BEST, Galaxy-BioProd

Phylogenetic Analysis of *Francisella tularensis* Outbreak Strains in Bulgaria Using a Reproducible Galaxy-Based Workflow

TOLCHKOV Vladimir, NENOVA Romyana, TOMOVA Iskra, IVANOV Ivan, DONCHEV Deyan, STOIKOV Ivan, HODZHEV Yordan, DOBRINOVA Yordanka, POCHILEEV Iliyan, MONEV Hristiyan, TSAFAROVA Borislava

Poster only

Francisella tularensis, the causative agent of tularemia, is a highly infectious zoonotic pathogen that has caused recurrent outbreaks in Bulgaria. Understanding the phylogenetic relationships between outbreak strains is essential for effective epidemiological surveillance and response.

In this study, we developed a reproducible workflow within the Galaxy platform for comparative phylogenetic analysis of *F. tularensis* strains. The dataset includes isolates from two distinct outbreaks in Bulgaria: West Bulgaria (2003) and Northeast Bulgaria (1962). Strains were obtained from the National Reference Laboratory for Special Bacterial Pathogens and include human, animal, and environmental isolates, as well as control and vaccine strains.

Whole-genome sequencing was performed using Illumina HiSeq technology. Raw reads (FASTQ format) were processed and assembled using the Unicycler tool integrated within Galaxy. The resulting assemblies were evaluated, and the genome with the fewest contigs was selected as the reference for downstream analysis. Phylogenetic reconstruction was conducted using the RealPhy platform.

The workflow revealed clear phylogenetic separation between outbreak groups, while isolates from the same outbreak source showed minimal genetic variation. These findings are consistent with previous analyses using AFLP and MALDI-TOF but provide substantially higher resolution.

Our Galaxy-based pipeline ensures reproducibility, accessibility, and scalability of whole-genome phylogenetic analyses. It can be readily adapted for routine use in public health laboratories and supports improved outbreak investigation and molecular epidemiology of high-risk bacterial pathogens.

Keywords: *Francisella tularensis*, Phylogenetic Analysis, Realphy

Acknowledgements: This work is supported by BG-175467353-2025-07-0061, the “Center of Competence Immuno-Pathogen”, procedure BG16RFPR002-1.014; “Sustainable Development of Centers of Excellence and Centers of Competence, including specific infrastructures or their associations of the NRRI”; and the “Research, Innovation and Digitalization for Smart Transformation 2021–2027” program

de.NBI – A National Infrastructure for Data-Driven Life Sciences

SRIKAKULAM Sanjay Kumar, FRENTRUP Martinique, LÜBKE Nils-Christian, JOHANNA Nelkner, SCHNITZER Helena, WIBBERG Daniel, MAUS Irena, SCZYRBA Alexander, KOHLBACHER Oliver, DAMMANN-KALINOWSKI Tanja, HOFFMANN Nils

Poster only

The German Network for Bioinformatics Infrastructure (de.NBI) is a national, academic, non-profit research infrastructure supporting data-driven life sciences in Germany and beyond. Funded at the federal level and coordinated within the Helmholtz Association, de.NBI connects 24 partner institutions that provide bioinformatics tools, services, and training for academia and industry.

de.NBI promotes FAIR data practices, reproducible workflows, and community-driven tool development, contributing to efficient and sustainable bioinformatics research. Through ELIXIR Germany and alignment with European initiatives such as ELIXIR Europe and the European Open Science Cloud (EOSC), de.NBI supports the development of standards in research data management, training, and interoperable infrastructure.

An important component of this ecosystem is the federated de.NBI Cloud, which provides free and scalable compute resources for large-scale data analysis, ML/AI analyses, virtual research environments, and the operation of research tools and web services. It supports widely used community platforms such as Galaxy Europe, enabling them to serve a large and international user base.

Current work extends the de.NBI Cloud towards more flexible compute integration for Galaxy Europe. Building on the existing SimpleVM infrastructure, this includes enabling Bring Your Own Compute (BYOC) through automated provisioning and Pulsar-based execution backends. Both fully managed and user-managed deployment models are supported, allowing seamless integration of additional compute resources while maintaining reproducibility and ease of use.

This poster highlights how de.NBI enables scalable and reproducible life science research through its integrated infrastructure, services, and ongoing development efforts.

The Biodiversity Genomics Galaxy Lab: A Collaborative Hub with curated tools, workflows and training

CORREARD Solenne, BATUT Bérénice, ELIXIR Biodiversity Community, WATERHOUSE Robert M., BRETAUDEAU Anthony

Poster only

Bioinformatics is a rapidly evolving field, with countless tools, workflows, and tutorials. For researchers, identifying relevant and up-to-date resources can be daunting. Galaxy, an open, web-based platform, is widely adopted for FAIR-compliant and reproducible bioinformatics analyses across many domains, including biodiversity genomics. To make Galaxy more accessible and less overwhelming, specialized “labs” can be created: public instances with curated tools, workflows, and tutorials tailored to specific scientific communities.

Biodiversity is an incredibly vast topic, spanning genetic diversity, species interactions, and ecosystem dynamics. Biodiversity genomics, a sub-domain of biodiversity research, is a very dynamic field, as shown by the numerous international and national initiatives such as the Earth BioGenome Project (EBP), European Reference Genome Atlas (ERGA), ATLasea (France), International Barcode of Life (iBOL), or Biodiversity Genomics Europe (BGE). These projects generate reference genomes and analyze genomic data for conservation, evolutionary biology, and the discovery of novel biomolecules, accelerating tool development and attracting a diverse range of stakeholders.

To address the need for accessible, high-quality bioinformatics resources to analyze biodiversity genomics data, we developed the Biodiversity Genomics Galaxy Lab. This user-focused platform organizes Galaxy’s extensive resources to meet the specific needs of biodiversity genomics researchers. The lab’s creation is a semi-automated process: relevant tools, workflows, and tutorials are automatically sourced from public repositories (i.e. Galaxy ToolShed, WorkflowHub, Galaxy Training Network) based on EDAM categories, keywords and tags. Then, manual curation is performed to exclude irrelevant or outdated content. The majority of the manual curation for the biodiversity genomics lab was performed by the Elixir Biodiversity Community during a workshop held at the ELIXIR ALL HANDS Meeting in 2025. Automatic queries of the public repositories of tools, workflows, and tutorials are performed weekly to ensure the list remains up-to-date and relevant to users. As of March 2026, the lab includes 106 tools, 42 tutorials, and 33 workflows.

The lab is largely based on tools, workflows, and tutorials produced by the Galaxy Ecology community and focuses on biodiversity genomics. Curated by the Elixir Biodiversity Community, it integrates only community-relevant features, simplifying navigation and enabling scientists to focus on advancing biodiversity genomics research. While the lab currently emphasizes genomics-based approaches, we recognize the importance of exploring all dimensions of biodiversity.

The lab is accessible at:

- Galaxy EU: <https://biodiversity-genomics.usegalaxy.eu>
- Galaxy FR: <https://biodiversity-genomics.usegalaxy.fr>
- Galaxy ORG: <https://biodiversity-genomics.usegalaxy.org>

Microbiology Galaxy Lab: The first community-driven gateway for reproducible and FAIR analysis of microbiological data

BATUT Bérénice, NASR Engy, PECHLIVANIS Nikos, STREPIS Nikolaos, ZIEREP Paul

Poster only

The exponential growth of microbial omics data has transformed microbiology, yet many researchers face barriers in analyzing these complex datasets due to limited computational resources and expertise. To address this challenge, we present the Microbiology Galaxy Lab (MGL), a globally accessible, community-driven platform that integrates cutting-edge bioinformatics tools with user-friendly accessibility.

Built on the Galaxy framework, MGL provides a comprehensive environment for analyzing (meta)genomic, (meta)transcriptomic, and (meta)proteomic data, aligning with FAIR principles to ensure reproducibility and interoperability. The platform offers over 315 specialized tools and 115 curated workflows, enabling researchers to perform advanced analyses such as antimicrobial resistance tracking, microbiome classification, and functional annotation of microbes. MGL supports a wide range of uses, from studying health and infectious diseases to exploring environmental microbiology, enabling researchers to connect data collection with valuable biological understanding.

A key strength of MGL is its emphasis on community engagement and education. The platform includes over 35 tutorials and learning paths, empowering researchers at all skill levels to adopt best practices in microbial data analysis. By fostering collaboration and standardization, MGL accelerates discovery and promotes interdisciplinary research, ultimately advancing our understanding of microbial roles in health, disease, and global ecosystems.

Available on multiple Galaxy servers (e.g., microbiology.usegalaxy.org, [.eu](https://microbiology.usegalaxy.eu), [.org.au](https://microbiology.usegalaxy.org.au), [.fr](https://microbiology.usegalaxy.fr)), MGL democratizes access to powerful bioinformatics tools, enabling researchers worldwide to overcome computational barriers and drive innovation in microbiology. This poster will highlight MGL's features, use cases, and its potential to transform microbial omics research through open, reproducible science.

From SAM Segmentation to YOLO Detection: Enabling Semi-Automatic Marine Biodiversity Annotation in Galaxy for Citizen Science Application

BARREAU Arthur, LE BRAS Yvan, LE BRIS Nadine, SEGUINEAU Pauline, GALAXY ECOLOGY Community

Poster only

The MOOREV project, launched in 2022, aims to improve understanding of the effects of climate disturbances on the seashore and its biodiversity. Built as a citizen science initiative, it brings together researchers, environmental education professionals, schools and local communities, using the shoreline as a natural laboratory. Its approach is based on deploying cameras in tidal pools, allowing local species to be observed at the individual level. Extracting scientific value from these data requires significant processing: identifying species, but also measuring behavioral and physical parameters such as size or displacement speed and direction. The volume of data collected makes this processing a central challenge for the project.

To address this, Galaxy was chosen for two main reasons. First, the YOLO Galaxy tools already available allow object detection models to be trained and used to automate image analysis. Second, the interactive tool AnyLabeling makes it possible to manually annotate images and apply segmentation models. However, manually segmenting and annotating hours of video is time-consuming. The recent release of SAM3 (Segment Anything Model 3) came at the right time: it introduces semantic input, allowing images and videos to be segmented from a simple text prompt. A dedicated Galaxy tool for SAM3 was developed, supporting both images and videos as input.

The objective of this work is to build a Galaxy workflow for automatic analysis of images and videos. Training data is first processed by SAM3, which outputs segmentation masks in COCO format, then imported into AnyLabeling where users can correct, add or remove annotations. The validated annotations are used to train a model with YOLO Training, which can then be applied with YOLO Predict to automatically analyze new content.

This workflow shows how Galaxy can support citizen science projects in marine ecology, and provides a reusable base for other species detection and annotation projects.

BioFAIR Pathfinder: Connecting Bioconductor, Galaxy, and nf-core: FAIR, AI-ready workflows and training for single-cell analysis

RUE-ALBRECHT Kevin, SIMS David, LOACH Marisa, DOYLE Maria, COOLEY Nicholas, RIOUALEN Claire

Poster only

Researchers analysing single-cell data face a fragmented tool landscape: many excellent methods exist across communities, but inconsistent packaging, documentation and metadata make it difficult to find, combine or reuse the right tools for a given biological question. **Bioconductor** provides a rich set of R-based software packages for single-cell analysis, while **Galaxy** offers a graphical, reproducible environment that makes these tools accessible to users without programming experience.

Here, we present the roadmap of a **BioFAIR Pathfinder** project that is working to improve how Bioconductor tools and training materials are shared across workflow platforms, making them FAIR-er and better prepared for future AI-driven applications.

The project brings together expertise from **Bioconductor**, **nf-core**, and the **ELIXIR** and **Galaxy** communities (including Galaxy, Single-Cell Omics, and SPOC), building on existing infrastructure such as **WorkflowHub**, **bio.tools**, **RDMkit**, and **EDAM**. It also aligns with the work of **BioFAIR Fellows** Marisa Loach and Nicola Soranzo. The project aims to measurably improve Bioconductor FAIR-ness and facilitate the creation and dissemination of exemplar **single-cell analysis** workflows that demonstrate how Bioconductor tools can be packaged, described, and registered in a way that supports reuse across platforms and prepares them for future AI-driven discovery.

Started in April 2026, this 12-month project aims to deliver:

1. Tooling for adapting Bioconductor methods as **Galaxy wrappers** and **nf-core** modules
2. Extensions to existing **Galaxy single-cell workflows** demonstrating interoperability with **Bioconductor** software
3. Guidance for alignment of Bioconductor metadata with **EDAM** and registration via **WorkflowHub** and **IWC**
4. Tutorials published through the **Galaxy Training Network (GTN)**, the **ELIXIR Training e-Support System (TeSS)** and its federated hub **mTeSS-X**, Bioconductor platforms and the forthcoming Bioconductor-JOSE (**Journal of Open Source Education**) partnership.
5. Contributions and improvements to the single-cell **RDMkit**

Facilitating the creation of galaxy tools

LE CRAS Jean, LE BRAS Yvan, SEGUINEAU Pauline, GALAXY ECOLOGY Community

Poster only

Creating Galaxy tools can be quite challenging for non-programmers, which is why methods have been developed to simplify the process, particularly using Jupyter Notebooks.

Nb2workflow (<https://gxy.io/GTN:T00567>) is a Python package developed by the astronomy Galaxy community that converts Jupyter Notebooks into workflows; within this package, nb2galaxy allows users to generate a Galaxy tool consisting of a Python script and a wrapper from a Jupyter Notebook with annotated code cells indicating which parameters and outputs are used. The purpose of the improved version of nb2galaxy is to support notebooks written in R; the current version only supports notebooks written in Python.

Some Galaxy tools may originate from other pipeline engines and have had to be adapted to meet Galaxy's technical constraints. In this context, the aim of this work is also to develop a library to convert Bon in a box tools into Galaxy tools.

The Biab-to-Galaxy tool automates part of the process of porting Bon in a box tools to Galaxy. It generates a Galaxy XML wrapper from a Biab YML wrapper and modifies the tool's script to run in Galaxy. Bon in a box is a pipeline engine developed by GEO BON that focuses on Earth data analysis and the production of essential biodiversity variables. This converter could provide Galaxy with new tools in these categories, it could also join Planemo in the future.

Updating a ten years old metabarcoding Galaxy toolsuite for Galaxy Ecology and applying to Antarctic conservation needs

PAYET Kevin, DURAND Morgane, GALLUT Cyril, LE BRAS Yvan, SEGUINEAU Pauline,
GALAXY ECOLOGY Community

Poster only

Environmental genomics is a fast-developing area, featuring many innovative methods to characterize and measure biodiversity. Metabarcoding is one of the most used methods in environmental DNA. It can be used in a wide range of contexts, such as establishing an inventory of various taxa in a fragile environment, detecting potential invasive species before they thrive, or gaining a better understanding of the composition of the microbiota in the human gut. This method offers numerous benefits, notably being the simplest approach for detecting taxa with minimal environmental disruption. For long-term monitoring and to facilitate cross-studies, the need to go FAIR (Findable, Accessible, Interoperable, Reusable) is a must to reach eDNA's full potential. Galaxy is based on three essential values for research: accessibility, reproducibility, and transparency. A first step toward a FAIR method in eDNA is to have access to a useful set of tools within Galaxy. To process metabarcoding datasets, one of the most common toolsuites is OBITools. It can handle metabarcoding datasets from the raw data to the taxonomic attribution. The implementation of this toolsuite into Galaxy seems to be a valuable addition to handle this type of dataset. A ten-year-old version of OBITools already exists in Galaxy and the purpose of this work is to provide an updated version and add new tools to have the full scale of the toolsuite. Indeed, OBITools is now in version 4, with a complete rewriting of the code in Go, initially written in Python. This new version of OBITools tends to be more versatile and more stable than the previous version. Advances in sequencing technologies now make it easier to generate multi-marker datasets. However, these datasets require specific workflows to be processed correctly. Finally, the aim of this work is to use the obitools Galaxy toolsuite to develop a workflow for a multimarker dataset from an Antarctica use case and its associated Galaxy Training Network tutorial. The data in this dataset was collected for a study aiming to characterize benthic cryptofauna and examine the ecological succession of this fauna by utilizing ARMS (Autonomous Reef Monitoring Structures).

Reusing 15 years old scientific work to guide Galaxy Ecology tools creation for marine spatial planning

MAYI Emilie-Jeanne, LE BRAS Yvan, ELEAUME Marc

Poster only

Reproducibility in scientific research is an increasing concern, as it depends on the ability to preserve and reuse data and previous analyses to assess progress. In this context, we investigate the reproducibility of legacy methods and the long-term preservation of scientific data.

We focus on the Marine Spatial Planning in East Antarctica. The 2014 CCAMLR report aimed at identifying and predicting Marine Protected Areas (MPAs) for the East Antarctica Planning Domain. This study relies on complex workflows that may become difficult to access and reuse over time, despite years of development.

Here, we leverage 15 years of legacy analyses to reconstruct MPAs-related workflows for East Antarctica, and transform them into reusable tools within the Galaxy Ecology platform. By standardizing and integrating existing R and geospatial workflows (GIS), we aim to improve their reproducibility, accessibility, and scalability. This work demonstrates how Galaxy enables the preservation, execution, and adaptation of legacy analyses in a modern computational environment.

Galaxy customisation at OSUG: flexible HPC integration & easier workflow migration

VISAN Vlad, BOURDON Vincent, CAILLETAUD Rémi, ROCH Françoise, ROGEZ Yves, THOLLARD Franck

Poster only

Context

OSUG (Observatoire des Sciences de l'Univers de Grenoble) is a federative research structure that includes laboratories in astrophysics, earth sciences and, ecology. These groups use intensive instrumental and simulation data processing in complex workflows. Its main HPC resource is provided by GRICAD, while project-specific servers are also used. Galaxy was selected for workflow management, thanks to its features: workflow design, monitoring including restarting and caching, dataset and log visualization, RBAC, replicability and extensible job runners.

Specifically, research teams used to develop workflows without a workflow-manager, on large data volumes (terabyte-scale). Moreover, the heterogeneity of the user-community forbids the storage and duplication of produced data on a centralized server.

This led to specific Galaxy adaptations. User-wise: non-expert users should be able to create and update tools with minimal effort, and have group-based sharing for them. Cluster-wise: no cluster-side Galaxy-related installation should occur, data/logs shouldn't be duplicated in Galaxy, and accounting on the external clusters should be seamlessly mapped to the Galaxy users.

Galaxy providing native replicability features, we also use a companion software, AMI (LPSC), for advanced data traceability, versioning and exploration.

Motivation

These new features could tailor the experience for existing users, making the tool usage more ergonomic and consistent with the rest of Galaxy (in-place editing and sharing). They also make Galaxy viable for those with similar external clusters and non-duplicated data needs that currently cannot use it, while also making it easier to install on any new clusters since we removed the need for admin access.

Approach

We customized Galaxy in multiple ways:

- Backend plugins: Object Store for “shallow” (non-Galaxy/Pulsar and non-job-workdir) data, REST+JWT Job Runners (Slurm/OAR) specialized for Pulsar-less cluster execution.
- Web plugins for our custom tool manager: endpoints (front, back), and new pages .
- Minimal codebase adjustments meant to be vanilla-compatible: visualization and vault.

- Tools: some adapters (ex: upload replacement) and custom fields for shallow data (automatic since we control tool-generation).
- Clusters: companion and setup scripts.
- Ansible: in-house deploying of a cluster module, and backing up and restoring of non-Shed-stored tools.

Results

The complete solution is currently under alpha-testing. We defined 4 scenarios involving 4 different thematic research groups for the validation of the overall features. A large effort has been invested to make the full solution maintainable for long-term usage. The required adaptations lead to some restrictions in usage of Galaxy : datasets / files linkage is weak and can be broken by external user, as the files are not managed by Galaxy itself, which makes the replicability less reliable; some of the embedded visualization tools are not yet available; the community tools are not directly usable.

Conclusions and further work

This work has shown it is possible to have a flexible, externalized HPC resource at the cost of some replicability and code maintenance. We also made it easier migrate existing workflows. We intend to eventually integrate some of these features into Galaxy for everyone's benefit and have already started talking with some community members.

Galaxy as an enabling tool for uncovering the dark immunopeptidome through deep learning extensions

GRIFFIN Tim, DO Katherine, MEHTA Subina, WAGNER Reid, JAGTAP Pratik

Poster only

Immunopeptidomics identifies peptide antigens that could be developed as immunotherapies for cancer and other diseases. We have developed two complementary, Galaxy-based bioinformatics pipelines for immunopeptidomics analysis: the immunopeptidogenomics (iPepGen) pipeline (Genome Biol, 2026, 27) and the related meta-immunopeptidomics (Meta-iPep) pipeline, designed to identify host neoantigen peptides and microbial peptide antigens, respectively. iPepGen uses host next generation sequencing (NGS) data to predict novel neoantigen peptide candidates, while Meta-iPep employs metagenomic information to define microbial peptide antigen candidates. Both pipelines match candidate peptides contained within customized protein sequence databases to tandem mass spectrometry (MS/MS) data and employ tools predicting peptide binding to human leukocyte antigen (HLA) complexes as an indicator of immunogenicity and therapeutic potential.

Although effective for discovering many peptide antigens, these pipelines have limitations. For iPepGen, RNA preparation artifacts, low expression levels and/or short transcript half-life, and other limitations of the NGS sequence assembly process limit comprehensive detection of transcripts for predicting neoantigen peptides and building customized protein databases. For Meta-iPep, metagenomic information is not always available, making for incomplete protein sequence databases used to identify microbial antigen sequences. Moreover, both pipelines rely on matching MS/MS spectra to large sequence databases, which limits sensitivity when maintaining acceptable false discovery rates. These limitations give rise to a “dark immunopeptidome” undetectable using traditional methods. Deep learning-based de novo peptide sequencing of raw MS/MS spectra offers a solution. This approach takes raw, uninterpreted peptide MS/MS spectra and determines a sequence match with no dependence on a protein sequence database. Meanwhile, for identified peptide candidates, new deep learning models offer prediction of immunogenicity when paired with HLA genotyping information, enabling more efficient prioritization of identified peptides for further study.

In this presentation, we will describe ongoing work extending our Galaxy based pipelines with deep learning tools, including Casanovo for de novo sequencing (Nat Commun, 2024, 15) and pVACtools for immunogenicity prediction (Cancer Immunol Res, 2020, 8). Initial analyses on representative cancer immunopeptidomics datasets demonstrate that de novo sequencing identifies dozens of novel peptide candidates that were missed by traditional sequence database searching using our pipelines, including sequences encoded by pathogenic bacteria. These complement verified candidates from traditional sequence database searching, increasing confident HLA-bound peptides by >2-fold for further prioritization. Meanwhile, deep learning tools in pVACtools offer an efficient means to further prioritize these candidates. Galaxy implementation is key to enabling these extensions. Galaxy’s flexibility facilitates integration of deep learning tools into existing iPepGen and Meta-iPep modules. The results are directly compatible with our peptide-centric verification module and offers a means to

rigorously evaluate de novo identified candidates, and output highly confident peptide sequences for further consideration and prioritization using the pVACtools. Finally, compatibility of Galaxy with graphical processing units (GPUs) computing is being explored to promote high throughput analysis and potentially customized training of models used by the deep learning tools. Taken together, this presentation will demonstrate the value of the Galaxy ecosystem as a powerful tool for uncovering the dark immunopeptidome and advancing immunotherapy discovery.

When Galaxy Meets FAIR Standards: Building Interoperable Virtual Research Environments Across Disciplines

GRELET Sylvain, DELAPORTE Pascal, KEUCHKERIAN Samuel

Poster only

Several national and international initiatives rely on the Galaxy ecosystem to develop Virtual Research Environments (VREs) that have proven effective in supporting reproducible and collaborative research.

In parallel, the Open Science movement has established a strong framework for data discovery, sharing, and reuse through widely adopted standards and best practices (e.g., OGC, W3C, RDA, INSPIRE). Bridging these FAIR data principles with operational research platforms remains a key challenge for many scientific domains.

In this presentation, we share insights, activities, and lessons learned from deploying interoperable VREs across multiple communities, including water, geology, and risk, at both national and international scales. Our work explores how Galaxy, combined with tools such as Jupyter Notebooks, can be effectively integrated within a broader ecosystem of platforms, standards, and services.

We address several complementary dimensions: the adoption of FAIR practices within scientific communities; the articulation between heterogeneous platforms and data infrastructures; the implementation of both established and emerging interoperability standards through open-source solutions; and the integration of domain-specific codes and expertise into shared, reproducible workflows.

Beyond technical aspects, we highlight the importance of supporting changes in scientific practices. This includes engaging domain experts, facilitating the transition toward collaborative and documented workflows, and developing training programs tailored to diverse user profiles.

We also present a series of end-to-end use cases already implemented, along with the methodology used to extend these workflows to new domains and communities. These examples demonstrate how Galaxy can act as a central component within a federated ecosystem, enabling seamless interaction between data access services (e.g., MAD), processing environments, and visualization tools.

We also highlight the need to implement FAIR processes and services in and from our platforms. We developed an OGC API process gateway to expose the Galaxy workflows following international standards (Ghost) and are working with the communities (including Data Terra and its data providers) to develop services catalogs.

This work illustrates how aligning Galaxy with FAIR principles and interoperability standards can foster more sustainable, scalable, and cross-disciplinary research infrastructures, while addressing both technical and organizational challenges.

The Ever Evolving Galaxy Workflow Graph: Editor and Invocation View Polish

AWAN Ahmed, VAN DEN BEEK Marius, NEKRUTENKO Anton, SCHATZ Michael

Poster only

Since the introduction of the Workflow Invocation View in GCC 2024, Galaxy's workflow graph canvas has seen many improvements on both the pre-run editor and the post-run live-updating invocation views. In this talk, we will highlight all the latest improvements to Galaxy's Workflow Graph, and how these features have grown around the work presented at recent GCCs.

As users edit and create their workflows, they often miss out on key best practices that can lead to more meaningful and reproducible workflows. Such practices include using formal input parameters, labeling workflow inputs and outputs, and attributing appropriate metadata to the workflow. We have, therefore, modernized the workflow editor's best practices panel which tracks these requirements, grouping them into critical and attribute best practices.

Apart from improvements that assist workflow creation within the editor, including the best practices panel and also minor polishes like tracking the undo/redo change stack counts, we have made improvements to the base graph canvas itself, used in both the editor and invocation views. A "pipeline-isolation" styling effect has been added to the graph which, when currently editing a step in the editor or observing the jobs at a step in the invocation view, dims all non-connected (non-upstream and non-downstream) nodes. This allows users to focus on the relevant steps and their connections, especially in large workflows.

The invocation view side has also seen extensive polish, including minor features like the ability to toggle the visibility of inputs and outputs directly on the graph (mimicking the current active step like it would have appeared in the editor), a debug tab allowing email reports for workflow runs, etc. Major polish has also been applied to the currently viewed step display, which now features separate Jobs, Outputs and Step Config tabs, a paginated jobs table, navigation buttons to jump between steps, and more.

Every bit of the extensive polish that has been done to the long running Galaxy workflow canvas has been driven by feedback, and we aim to continue improving such core interfaces via input from the community. We look forward to sharing these features at GCC 2026 and gathering the feedback that will drive what comes next.

Using a Galaxy Workflow in an ISO-Accredited lab

VAN HEUSDEN Peter, BUHLE Ntozini

Poster only

ISO standards such as ISO 15189 and ISO 17025 provide essential guidance for quality management in medical testing laboratories, including those serving public health. Such standards and the accreditation processes used in their use are, however, often foreign to bioinformatics workflow authors. Using standard operating procedure (SOP) templates, truth sets and workflow validation illustrated using a Galaxy workflow for *M. tuberculosis* sequence analysis, the AdvISO project is developing a toolkit for quality-assured pathogen bioinformatics. We present some of the challenges and learnings from translating ISO 15189 and 17025 concerns and procedures to a bioinformatics context.

User-friendly genome assembly and annotation workflow on Galaxy platform to track foodborne pathogens in novel seaweed-based food

RANDAZZO Walter, ALCÁINE BELLIDO Beatriz, GARCÍA GUTIÉRREZ Enriqueta, BAKER David, NARBAD Arjan, RODRIGO Dolores

Poster only

In foodborne outbreak investigations, whole-genome sequencing (WGS) is widely used to address key questions such as species identification, detection of antimicrobial resistance genes (ARGs) and virulence factors, and source attribution. The availability of user-friendly, automated, and interlaboratory-validated workflows enables the rapid implementation of mitigation strategies along the food chain, which is particularly important for safeguarding public health and consumer safety. This is especially relevant for novel food products involved in global trade, such as edible seaweed. In this study, we implemented a workflow on the Galaxy platform for the WGS assembly and annotation of Illumina short-reads from different bacterial species isolated from various seaweed species, enabling reliable taxonomic identification and risk assessment.

Dispatching Galaxy Jobs to Google Batch: Direct API vs Pulsar AMQP Sidecar

SUDERMAN Keith, CHILTON John, AFGAN Enis, NEKRUTENKO Anton, SCHATZ Michael

Poster only

Galaxy deployments on Kubernetes typically execute jobs within the cluster, constraining scalability to available cluster capacity. Newly added support for Google Cloud Batch offers an alternative model in which jobs run on dynamically provisioned VMs with per-tool resource sizing, eliminating cluster capacity constraints.

We implemented two job runners for dispatching Galaxy tools to Google Cloud Batch. The first runner, `GoogleCloudBatchJobRunner` (Direct), submits jobs directly to the GCP Batch API, staging data via an NFS volume shared between the Kubernetes cluster and the Batch VMs. The second runner, `PulsarGcpBatchJobRunner` (Pulsar), dispatches jobs through a Pulsar sidecar container on each Batch VM, with Galaxy communicating via AMQP (RabbitMQ) and Pulsar handling file transfer via HTTP to a local SSD. We also evaluated hybrid configurations (Direct+K8s and Pulsar+K8s) that route lightweight single-core tools; parameter tools, text processors, and simple BAM/BED utilities; to the local Kubernetes runner, reserving GCP Batch for compute-intensive tools.

Both runners integrate with Galaxy's Total Perspective Vortex (TPV) routing system to select the appropriate machine size based on each tool's CPU and memory requirements. Custom VM images with pre-installed CVMFS clients provide immediate access to reference data.

We performed a cost and performance analysis across three IWC workflows: variant calling (12 tools), RNA-seq (12 tools, 20 steps), and ChIP-seq (7 tools) at three dataset sizes (2 GB, 5 GB, and 10 GB). Wallclock-based billing reveals significant differences between configurations. Plain Pulsar is the most expensive at \$41.39 total, due to per-job local SSD provisioning and AMQP staging overhead on every tool. Plain Direct costs \$32.60, with NFS-based staging introducing less per-step overhead. The hybrid configurations substantially reduce costs by eliminating GCP Batch VM provisioning for lightweight tools: Direct+K8s costs \$27.82 (15% less than Direct) and Pulsar+K8s costs \$32.99 (20% less than Pulsar). Compute-only costs (cgroups runtime) are nearly identical across all runners (\$3–4 per workflow), confirming that cost differences are entirely attributable to infrastructure overhead: VM provisioning, container image pull, and file staging.

The overhead ratio — wallclock cost divided by compute cost — ranges from 4–5x for Direct to 5–6x for Pulsar. Routing lightweight tools to the Kubernetes runner reduces this overhead by 15–31%, with

the largest gains on workflows containing many parameter tools such as RNA-seq.

These results suggest that the direct Batch runner with hybrid Kubernetes routing for lightweight tools is the better default for multi-step workflows, offering the best balance of cost and simplicity. The Pulsar-based runner remains valuable for deployments without a shared filesystem and for long-running, I/O-intensive single-step jobs where staging overhead is amortized and local SSD provides measurable performance gains. Future work includes replacing NFS with a GCS object store backend to improve throughput for I/O-intensive jobs, support for spot/preemptible VMs, and automatic per-tool runner selection based on I/O and resource characteristics.

Towards Democratisation of Whole Genome Alignments

SEACORD Andrew

Poster only

More genomes are being sequenced today than ever before, yet only select groups of researchers possess the computational hardware and technological know-how to turn DNA assemblies into scientific advancements. Leveraging Galaxy's user-friendliness, computing power, and overall practicality, we created a highly versatile workflow to ease this common gap of post-assembly processing. The fundamental core of this workflow is to simplify and speed-up the conversion of DNA sequences to multiple genome alignments. Utilizing a minhasher algorithm to estimate pairwise distances for the input set of genomes, the distance matrix will sort each pair based on a threshold to be aligned with one of two pairwise aligners: FastGA for similar sequences, and KegAlign (a GPU-accelerated fork of LastZ) for more divergent sequences. The distance data will also be used to construct a phylogenetic tree. This tree, combined with the complete set of pairwise alignments, will then be fed into Multiz to generate a multiple genome alignment. A process that often takes hours of time from researchers to download software, manipulate file formats, and learn software-specific commands/options, will be distilled down into a Galaxy workflow that works right out of the box, is well-documented, and has better performance. This workflow will still provide access to all options and features and will run in merely a fraction of the time without all the headache. Alignments are absolutely essential for both understanding fundamental biology and conducting clinical & biomedical research. This new workflow, built on cutting edge tools, will expand the opportunities for researchers to benefit from the many promising resources of comparative genomics.

Leveraging galaxy tools for clinical variant processing and interpretation in inherited retinal diseases

LÓPEZ-RODRÍGUEZ Víctor Rubén de Jesús

Poster only

Inherited Retinal Diseases (IRDs) present significant diagnostic challenges due to their genetic heterogeneity. While whole exome sequencing is a cornerstone of clinical diagnostics, the transition from raw data to a molecular diagnosis remains a challenge for clinicians. This work details the implementation of a bioinformatic pipeline using the Galaxy platform for IRD variant processing and interpretation. By leveraging Galaxy's interface, we integrated tools for quality control, alignment, and variant calling. This workflow emphasizes reproducibility and clinical relevance, allowing for the systematic identification of causative mutations in rare retinal disorders. The use of Galaxy bridges the gap between bioinformatic complexity and clinical application, empowering medical geneticists to enhance diagnostic yield for patients with IRDs.

Genome Analysis Suggests Reclassification of a Putative *Clostridium botulinum* Strain

GARCIA Priscilla, VAN LAAR Tricia

Poster only

Accurate identification of toxigenic *Clostridium* strains is important for both public health and microbial classification. We analyzed Illumina sequencing data (SRR29831631), initially labeled as *Clostridium botulinum*, to evaluate toxin gene content and taxonomic placement using a reproducible Galaxy-based workflow.

Reads were quality assessed using FastQC and filtered with fastp (Q30 >97%). The genome was assembled using SPAdes and evaluated with QUAST, yielding a high-quality draft assembly (~4.0 Mb, N50 ≈ 746 kb, 30 scaffolds). Genome quality was further assessed using CheckM2, indicating high completeness (99.99%) and low contamination (1.66%). Annotation using Prokka identified genes associated with sporulation, motility, and mobile genetic elements.

To evaluate toxigenic potential, botulinum neurotoxin genes (*bont*) were investigated using Prokka annotation, ABRicate, and targeted BLAST searches. No complete *bont* genes were detected, and only short alignments (<100 bp) were observed, suggesting the absence of a functional toxin gene cluster.

Ribosomal RNA genes were identified using Barrnap, and the 16S rRNA sequence showed high similarity to both *Clostridium botulinum* and *Clostridium sporogenes*, highlighting the limited resolution of 16S-based identification within this group.

Taxonomic identity was further resolved using Average Nucleotide Identity (ANI) analysis with fastANI against a curated reference set. The genome exhibited ANI values of 95–97% with *C. sporogenes* and ~92–93% with *C. botulinum*. These results support classification within *Clostridium sporogenes*, consistent with clustering observed in ANI heatmaps.

Overall, this study demonstrates that the analyzed strain is non-toxigenic and more consistent with *Clostridium sporogenes* than *Clostridium botulinum*. This project shows the importance of genome-wide analysis over single-gene approaches and demonstrates how Galaxy-based workflows can be used to resolve taxonomic ambiguity in closely related bacterial groups.

Pan-Disease Transcriptomic Signatures for Translational Target Discovery

KOIRALA Shubham, ZHANG Xiaodan, GURAGAIN Aastha, JOSEPH Dimitri, PALMER Zolt, SHANKAR Rama, LESHCHINER Dmitry, CHEN Bin

Poster only

Precision medicine seeks to redefine human disease based on molecular mechanisms rather than clinical symptoms, yet transcriptomic data—one of the most abundant and informative molecular modalities—remain fragmented, inconsistently annotated, and difficult to integrate across studies. To address this, we developed a scalable computational framework to construct a pan-disease reference atlas of transcriptomic signatures from large-scale public RNA-seq data.

We aggregated and uniformly processed bulk RNA-seq datasets and standardized heterogeneous sample metadata using an AI-based annotation framework. We then generated robust disease-specific signatures using data-driven surrogate control selection for studies lacking matched controls, coupled with cross-study meta-analysis. This approach enables systematic cross-disease comparisons and supports translational applications such as therapeutic discovery through integration with perturbational datasets.

To improve biological interpretability, we characterized the disease immunome by performing large-scale immune deconvolution of bulk transcriptomic profiles. By integrating multi-tissue single-cell RNA-seq-derived immune reference datasets, this framework disentangles cell-intrinsic transcriptional changes from shifts in immune cell composition, enabling immune-aware analysis across diverse disease contexts.

As a translational application, we applied this framework to cardiogenic shock, integrating bulk RNA-seq, single-cell RNA-seq, and plasma cytokine data from a clinically well-phenotyped cohort. This analysis identified immune cell states and transcriptional programs associated with disease severity, providing mechanistic insight into immune dysregulation in critical illness.

We are actively translating this framework into modular, reproducible workflows within the Galaxy ecosystem, with the goal of enabling accessible, large-scale transcriptomic integration and analysis. This work highlights both the challenges and design considerations involved in adapting complex, AI-assisted and cross-study meta-analysis pipelines for Galaxy, and aims to contribute reusable tools and workflows back to the community.

Implementing SensiTyper in Galaxy for personalized gonorrhoea treatment

GOLPARIAN Daniel, JAKOBSSON Jenny, LEONOR Sánchez-Buso, UNEMO Magnus

Poster only

Neisseria gonorrhoeae remains a major public health threat because of its extraordinary ability to acquire antimicrobial resistance, which increasingly limits empirical treatment options. Genome-based prediction of antimicrobial susceptibility is a promising strategy to support more individualized therapy, but broader use depends on workflows that are reproducible, accessible, and practical to deploy.

We describe our ongoing implementation of SensiTyper, a curated rule-based method for prediction of antimicrobial susceptibility in *N. gonorrhoeae*, within the Galaxy platform. The aim is to translate an expert-driven genomic interpretation framework into a transparent and user-friendly workflow that can be run, inspected, and shared more easily across laboratories. The workflow is designed to take genomic data as input and generate interpretable susceptibility-associated outputs relevant for treatment support and surveillance.

Implementing SensiTyper in Galaxy provides several advantages, including improved reproducibility, easier workflow management, reduced dependence on command-line expertise, and better opportunities for standardization and collaborative use. This is particularly relevant for clinical and public health settings where adoption of genomic methods may otherwise be limited by technical barriers.

This work represents a step toward making genome-informed gonorrhoea treatment support more accessible in routine laboratory environments. It also demonstrates the value of Galaxy as a platform for deploying clinically relevant bacterial genomics workflows in a reproducible and scalable way.

Multiomic Integration and Deep Learning–Driven Network Modeling for Reproducible Cancer Biomarker Discovery

ATTIQUE Zarlish, ANSARI Adnan Ahmed

Poster only

High-throughput transcriptomic and multiomic analyses generate complex, multi-layered datasets that require structured and reproducible computational workflows for meaningful interpretation. Integrating multiple data modalities with network-based analysis and predictive modeling remains a key challenge in translational bioinformatics, particularly when aiming to maintain consistency, scalability, and interpretability across studies. This work presents integrative analysis pipelines developed for transcriptomic and multiomic cancer datasets and their application within a reproducible workflow framework aligned with the Galaxy ecosystem. These pipelines extend standard RNA-seq processing by incorporating additional analytical layers, including gene regulatory network (GRN) reconstruction, protein–protein interaction (PPI) analysis, pathway enrichment, and hub gene identification. This enables a transition from single-layer analysis toward systems-level interpretation of disease mechanisms. The analytical strategies were applied across multiple cancer contexts to demonstrate their relevance. In pancreatic cancer, integration of transcriptomic, mutational, and epigenetic data combined with deep learning–based modeling enabled accurate patient stratification (AUC = 0.94) and identification of regulatory features associated with disease progression and survival outcomes. In metastatic breast cancer, transcriptomic and network-based analyses identified biologically significant driver genes and pathways, supporting downstream therapeutic prioritization through computational drug discovery approaches. Within this study, Galaxy-compatible workflow principles were used to structure preprocessing and analytical stages into reproducible and modular components. Standardized steps such as quality control, alignment, quantification, and differential expression analysis were organized to ensure consistency across datasets, while downstream analytical layers, including network modeling and predictive analysis, were incorporated as extended components within the overall workflow design. The study also highlights practical considerations encountered when applying reproducible workflow structures to complex, multi-layer analyses, including handling heterogeneous data types, maintaining parameter consistency, and integrating external computational environments. Emphasis is placed on maintaining transparency and reproducibility while supporting advanced analytical approaches. Overall, this work demonstrates the application of structured, reproducible workflows to integrative multiomic analysis and network-based interpretation, with alignment to the Galaxy ecosystem enabling accessibility, consistency, and scalability in biomarker discovery and translational research.

When and Where Proteins Move: RMSX and Flipbook for Molecular Dynamics

BERULDSEN Finn, ANTUNES Dinler

Poster only

Molecular dynamics (MD) simulations produce rich structural datasets, yet identifying when and where proteins move remains difficult and often requires specialized scripting. We introduce RMSX, a time-resolved extension of root-mean-square fluctuation (RMSF) that partitions trajectories to reveal when residue-level motions occur. To visualize RMSX and other residue-level features, we developed Flipbook, a visualization framework that maps time-series simulation metrics onto structural snapshots, producing colored and scaled representations that highlight precisely when and where conformational changes occur. Together, RMSX and Flipbook create intuitive and informative visual summaries of protein dynamics. This work addresses an underrepresented area within the Galaxy ecosystem. RMSX/Flipbook is open-source on GitHub, and ongoing work focuses on integrating these tools into Galaxy workflows for accessible and reproducible analysis of MD trajectories.

A Reproducible Galaxy Workflow for Genome Assembly and SNP Analysis of Hibiscus Chlorotic Ringspot Virus (HRS2 Isolate)

NIKITA Nikita, DHIR Sunny

Poster only

Reproducible and accessible workflows are essential for reliable analysis of high-throughput sequencing (HTS) data in virology. The Galaxy platform provides a user-friendly environment to address this need. Here, we present a Galaxy-based workflow for genome assembly and SNP analysis of Hibiscus chlorotic ringspot virus (HCRSV; species *Betacarmovirus hibisci*) from *Hibiscus rosasinensis*.

Raw HTS data were processed on Galaxy Europe. Read quality was assessed using FastQC, followed by trimming and filtering with Trimmomatic. High-quality reads were assembled de novo using Trinity, and contigs were screened against viral databases using NCBI BLAST+ (megaBLAST), leading to the identification of HCRSV with a genome size of 3.913 kb.

For downstream analysis, reads were aligned to the reference genome using Bowtie2, and variant calling was performed using DeepVariant. A total of 233 SNPs were identified, including 217 in coding regions (146 transitions and 71 transversions) and 16 in non-coding regions, with a transition/transversion ratio of 2.056. Among the total 293 protein-coding mutations, 207 were synonymous and 86 non-synonymous substitutions, with 39 conservative and 47 non-conservative changes, suggesting potential functional impacts.

The HRS2 isolate showed 92.63% nucleotide identity with the reference genome of HCRSV. This study demonstrates how Galaxy enables integrated, reproducible analysis of viral genomes from raw data to variant interpretation. The workflow can be readily adapted for similar studies in plant virus research and genomics.

Reimagining Uploads in Galaxy

LOPEZ TABERNERO David, HEIDARI Alireza

Poster only

This talk introduces Galaxy's new Beta Upload Activity, a redesigned data-ingestion experience that moves uploads from a modal dialog to a dedicated activity panel with a method-first workflow.

The new interface improves discoverability by presenting clear import paths (local files, URLs/paste, remote sources, libraries, archives, workflows, histories, and advanced rule-based import), each with contextual guidance. It also adds a staging area that lets users build batches incrementally, review items, and configure metadata before submission.

We will highlight practical workflow gains: integrated collection creation with validation and pair detection, destination history selection before upload, and a dedicated progress view for batch and per-item status tracking.

ABCP_finder: Transformer Embedding-Based Prediction of Anti-Breast Cancer Peptides

TRIPATHI Prabhat

Poster only

Anti-breast cancer peptide (ABCP) is a promising therapeutic class due to its high selectivity, low toxicity, and ability to target breast cancer specifically. However, the experimental screening of potential peptides is both time-intensive and cost-intensive. We present ABCP_finder, which is a transformer embedding-based computational framework for accurate identification of ABCPs. This method utilizes two pretrained protein language models (ProtBERT and ESM2). These two pretrained transformers encode amino acid sequences into higher-dimensional contextual embeddings to get structural and functional information. These embeddings are then combined with a range of machine learning models, including logistic regression, support vector machines, tree-based ensembles, and multilayer perceptrons. All classification models are brought together with a majority-voting ensemble strategy to enhance robustness. Across both independent test and validation datasets, all models delivered very strong performance with accuracy above 0.90, Matthew's correlation coefficient (MCC) around 0.82, and area under the curve (AUC) exceeding 0.96, demonstrating stable generalization and top-notch discriminative power. Embeddings derived from ProtBERT, when integrated with a multilayer perceptron, yielded the most optimal and equitable performance, but features based on ESM2, in conjunction with an ensemble model, exhibited strong classification capabilities. Final validation with activity prediction with previously unseen peptides showed clear separation in predicted potency. These findings support the biological relevance of the learned representations. ABCP_finder can provide an efficient, scalable, and safe screening platform that reduces experimental time and cost and accelerates rational discovery of ABCPs.

Administering a Galaxy Instance Tailored for Interactive Workshops

MAHMOUD Alexandru

Poster only

Galaxy has been the launch engine for Bioconductor Workshops for the past 3 years, leveraging Galaxy's Interactive Tools to give on-demand RStudio instances to Bioconductor users. This setup, hosted on Kubernetes, and managed by a single administrator, has had large success, especially in terms of ease of use for both users and workshop authors. The latter have benefited from an automated pipeline generating both workshop-specific containers and simple Galaxy wrappers which get automatically applied to the live instance, from workshops authored via the familiar BuildABiocWorkshop framework. This presentation will showcase these integrations, highlighting the latest improvements to the Galaxy-based workshop.bioconductor.org platform since the joint conference last year.

Unlocking Spatial Biology: End-to-End FAIR Workflows for Spatial Data Analysis in Galaxy

NAGHSH NILCHI Amirhossein, JUMAH Khaled, VIDEM Pavankumar, SPATIAL2GALAXY Consortium, GRUENING Bjoern

Poster only

Spatial omics technologies are transforming our understanding of cellular organization and tissue microenvironments, yet the volume, complexity, and multi-modal nature of spatial data remain significant barriers to accessible and reproducible analysis. Here, we present a comprehensive spatial data analysis ecosystem integrated into the Galaxy platform, enabling researchers to build robust, scalable, and FAIR-compliant workflows without requiring advanced programming expertise.

At the core of this ecosystem is the Galaxy SpatialData toolsuite, which provides a unified framework for spatial data handling. The `spatialdata-io` tool can convert the data generated from various spatial technologies, including Xenium, Visium, and MERSCOPE, into a standardized SpatialData object, while the `spatialdata-operations` supports complex spatial queries, multi-modal alignments, and data transformations. From the SpatialData suite, `spatialdata-plot` enables users to generate high-quality, customizable images from the SpatialData object from different layers or objects, colored by multiple annotations. For users who prefer the R ecosystem, we have also implemented Seurat spatial data visualizations.

For MERSCOPE-specific preprocessing, we incorporate the Vizgen Post-processing Tool (VPT), which handles transcript coordinate processing and cell boundary segmentation prior to SpatialData conversion. Multiple segmentation tasks using Cellpose, using pre-defined models or user-generated models, can be used to calculate an accurate cell segmentation and transcript allocation.

For downstream analysis, we have integrated SquidPy for spatial correlation and neighborhood graph construction, and LIANA+ for comprehensive cell-cell communication modeling, both accessible directly within the Galaxy interface and compatible with the SpatialData framework. For interactive data exploration, `spatialdata-napari` provides a dynamic visualization environment, allowing researchers to inspect and annotate spatial datasets without leaving the pipeline.

This SpatialData ecosystem is designed to be fully compatible with the existing Galaxy Single-Cell Lab, enabling users to seamlessly use the established single-cell tools and workflows in combination with newly integrated spatial tool suites. To showcase this, we developed Galaxy workflows inspired by popular spatial analysis workflows such as EISTA and WebAtlas from the Nextflow ecosystem, further improving them through enhanced integration of SpatialData and use of alternative, up-to-date tool suites. By adhering to FAIR (Findable, Accessible, Interoperable, and Reusable) principles, these workflows enable easy execution, reuse, and adaptation by users.

AI ecosystem for life science data analysis in Galaxy

KUMAR Anup, SALIKO Aerda, RIETHMUELLER Luca, KADKHODAEI Arash, NILCHI Amirhossein Naghsh, CHIANG Diana, GRÜNING Björn

Poster only

The rapid evolution of machine learning (ML) and foundation models has significantly transformed biological data analysis, yet their accessibility and reproducibility remain key challenges for many researchers. The Galaxy platform addresses these challenges by providing a web-based, FAIR-compliant environment that integrates advanced ML tools, compute infrastructure, and reproducible workflows into a unified interface. In addition to multiple ML tools categorised under the Galaxy-ML tool suite, new tools are being developed based on powerful foundation models trained on large-scale genomic sequences. These novel tools will work together with classical ML and deep learning (DL) tools to create end-to-end AI workflows for downstream tasks such as DNA and protein sequence classification, or regulatory element prediction. The integration of DNABERT-2 as a Galaxy tool will enable DNA sequence modeling in two modes - inference and fine-tuning on DNA sequences workflows directly within Galaxy, allowing users to leverage pretrained genomic embeddings for downstream tasks. Importantly, the workflows using the DNABERT-2 tool will be coupled with CNN-based tools available in Galaxy-ML, enabling hybrid modeling strategies. In such workflows, DNABERT-2 will serve as a feature extractor, generating contextual embeddings that are subsequently processed by a CNN architecture for improved task-specific performance. Fine-tuning pipelines will further allow users to adapt DNABERT-2 to domain-specific datasets, while maintaining reproducibility through Galaxy workflow sharing. Protein foundation models (such as ProtBERT and ProtTrans) will also be integrated as tools and then into Galaxy workflows. Similar to the DNA-based pipeline, these protein model-based tools will also support both inference and fine-tuning, enabling tasks such as protein function prediction and structure-aware classification. By combining these models with CNN-based architectures in Galaxy, users can design flexible pipelines where pretrained embeddings are refined through supervised DL. These tools, based on foundation models, will make advanced DNA and protein modeling accessible without requiring extensive coding expertise. Furthermore, to enhance usability and knowledge integration of LLM-based tools (such as LLM Hub), a retrieval-augmented generation (RAG) based tool has been integrated for extracting intelligent context from a large corpus and using it together with the LLM Hub tool for higher-quality LLM responses. This tool allows users to query external knowledge (such as scientific literature), enabling informed context selection, which is reflected in the better quality response by the LLM Hub. The existing AI ecosystem in Galaxy includes tools such as Flexynesis for multi-omics integration, TabPFN for efficient tabular learning, Cellpose for AI-powered image analysis, GPU-enabled JupyterLab for accelerated AI model development using GPU, and Galaxy-ML tool suite for classical ML and DL tools. These tools collectively support a wide range of AI applications involving varied datasets such as DNA, proteins, images, gene expression, natural languages, and biological networks. In summary, the AI ecosystem involving DNABERT-2, protein foundation models, CNN-based workflows, RAG-based tools, and specialized ML tools in Galaxy presents a comprehensive, reproducible, and user-friendly

platform for biological machine learning while lowering the barrier for adopting advanced AI methods in life sciences and ensuring reproducibility.

Genetic structure of *Aedes albopictus* in the Democratic Republic of the Congo

VULU Fabien, BOBANGA Thierry, MAMPUYA Pitshou

Poster only

Aedes albopictus, a major vector of dengue, Zika, and chikungunya viruses, has recently invaded the Democratic Republic of the Congo (DRC). Our previous analyses based on the mitochondrial COI gene revealed moderate haplotype diversity but low nucleotide diversity among populations from multiple regions, suggesting a recent introduction. To further investigate the population genomics of this invasive vector, we are conducting genome-wide analyses using targeted sequencing. Three *Ae. albopictus* specimens from each of 16 sites across the DRC were collected and processed for Illumina sequencing. Genomic DNA was fragmented, indexed, and enriched for approximately 200 genomic loci using a capture-probe approach developed by the Japanese National Institute of Infectious Diseases before sequencing on an Illumina NextSeq platform. Current analyses are being conducted using the Galaxy platform to perform quality control, read mapping, variant detection, and population genomic analyses. These results will provide the first genome-scale insights into the introduction history, genetic structure, and spread of *Ae. albopictus* in the DRC.

Galaxy@Pasteur: A Centralized Execution Engine for Bioinformatics Services

MAREUIL Fabien, PLANEL Rémi

Poster only

Galaxy@Pasteur¹, maintained at the Institut Pasteur, is widely used, with over 3200 active users, more than 670 available tools, and more than one hundred thousand jobs executed per month. Our platform supports two types of usage: interactive use through a web interface and automated use through the Galaxy APIs, illustrating its flexibility and strong adoption across a wide range of scientific and technical contexts.

In addition to its classic use, Galaxy@Pasteur stands out for its role as a centralized execution engine for bioinformatics web services. Beyond the user interface, it functions as a shared backend capable of orchestrating and executing complex workflows for third-party applications.

Our Galaxy platform is based on the computing infrastructure of the Institut Pasteur. This enables large-scale pipeline execution while simplifying access for client applications. Eight specialized web services (e.g., metagenomics, phylogeny, chemoinformatics...) interact with Galaxy² through its API to submit analyses, monitor their execution, and retrieve results. Galaxy thus ensures reproducibility, traceability, and dependency management.

To facilitate these integrations, the Institut Pasteur and the French Institute of Bioinformatics (IFB) developed and maintain django-to-galaxy³, an abstraction layer that simplifies interactions between Django-based web applications and Galaxy servers. This Django module provides features for Galaxy platform connection management, user configuration, job submission, parameter handling, and execution tracking, reducing application-side complexity and accelerating the development of interconnected services. It has notably been used in two projects: the ABRomics⁴ project, a national platform supporting antibiotic resistance monitoring and analysis, and DefenseFinder^{5,6} a web application designed to automatically detect and annotate known anti-phage defense systems in bacterial and archaeal genomes from nucleotide or protein sequences.

In parallel, a more generic approach called Galaxy-as-a-Service (GaaS)⁷ is under development. It aims to standardize the exposure of Galaxy as a service and facilitate its integration into distributed architectures. A dedicated poster presents this project.

Overall, Galaxy@Pasteur illustrates the evolution of a Galaxy instance into a service-oriented platform, central to a distributed application ecosystem and capable of supporting both interactive users and automated applications.

¹ <https://galaxy.pasteur.fr>

² The Galaxy Community. "The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update." *Nucleic Acids Research*, 2024, 52(W1):W83-W94. doi:10.1093/nar/gkae410

³ <https://gitlab.com/ifb-elixirfr/abromics/django-to-galaxy>

⁴ <https://www.abromics.fr/>

⁵ Tesson, F., Hervé, A., Mordret, E. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Commun* 13, 2561 (2022). doi:10.1038/s41467-022-30269-9

⁶ <https://defensefinder.mdmlab.fr/>

⁷ <https://github.com/rplanel/gaas>

Making a non-filesystem scientific resource browsable in Galaxy: the MaveDB example

POLUNINA Polina

Poster only

Researchers often require their data to be analysed in the context of external data sources, or simply want to re-analyse public data. To be most effective, scientific analysis platforms need to support data stored outside traditional filesystems. These external resources may be local or remote filesystems, require different network protocols, cloud object stores, repository platforms, or domain-specific scientific databases. Python's fsspec library addresses this diversity by providing a common filesystem-like interface over many storage and data access systems. The backend does not need to be a real filesystem; it only needs to expose useful filesystem-like operations such as listing resources, opening and reading file-like objects, and creating directories or writing files.

MaveDB is an open repository for Multiplexed Assays of Variant Effect datasets. It stores data sets containing processed scores, count data, variant tables, metadata, and mapped variant annotations. These data products are naturally useful as files, but MaveDB itself is organized around API concepts such as score set searches, public datasets, and authenticated user-associated datasets. Accessing many files across many score sets can require manual browsing, repeated downloads, uploads into Galaxy, or custom scripts.

We developed mavedb-fsspec, a fsspec adapter that exposes MaveDB through the mavedb:// protocol. Public score sets are available under mavedb://score-sets/, authenticated user-associated score sets under mavedb://my-score-sets/, and each score set appears as a directory containing files such as scores.csv, counts.csv, variants.csv, metadata.json, and mapped-variants.json when available. This mapping turns MaveDB's domain model into a stable virtual path hierarchy that can be used by Python tools expecting an fsspec-compatible filesystem.

We then integrated this adapter into Galaxy using Galaxy's Files Source framework. Files Sources provide a common way to expose external repositories in Galaxy, supporting repository browsing, user or administrator configuration, credentials, imports into histories, and exports for writable sources. The MaveDB Files Source subclasses Galaxy's shared fsspec-based implementation and opens the mavedb-fsspec filesystem, making MaveDB score set files available through Galaxy's repository browser and upload interface.

This approach improves on manual web or API-based access by allowing users to select one, several, or many MaveDB files directly inside Galaxy. Bulk import can support machine learning training datasets, benchmarking variant effect prediction tools, meta-analysis across assays, batch quality control, teaching datasets, and reproducible workflows that run across many score sets. The MaveDB example shows that even a domain-specific repository that is not inherently filesystem-shaped can become a practical, browseable Galaxy files source. More broadly, this work demonstrates a reusable

pattern for connecting scientific repositories to Galaxy: represent the resource with an fsspec filesystem, then expose it through Galaxy's Files Source layer.

Developing Metagenomic Workflows using Galaxy for the National Antibiotic Resistance Surveillance Platform (ABRomics-analysis)

LEFEUVRE Hugo, BATUT Bérénice, DIEUAIDE Amanda, SIGURET Cléa, MAREUIL Fabien, LAO Julie, TACKX Raphaël, VELO-SUAREZ Lourdes, HERY-ARNAUD Geneviève, PONS Nicolas, MEDIGUE Claudine, OLIVIER LECLERCQ Sébastien, RUPPÉ Étienne, HAYER Juliette, GLASER Philippe, BIHOUEE Audrey, CHAFFRON Samuel

Poster only

Abstract

Antibiotic resistance (ABR) is a critical global public health challenge, recognized as an urgent priority by international health institutions. The emergence and worldwide dissemination of multidrug-resistant (MDR) bacteria and antibiotic resistance genes (ARGs) are of particular concern, given their capacity to spread widely across humans, animals, and the environment. A dynamic that underscores the need for integrated One Health surveillance approaches.

ABRomics-analysis platform is an online service designed to track MDR bacteria and ARGs in a One Health context, from bacterial genomes and metagenomes. Its workflows are built on the Galaxy framework and have been validated by the Intergalactic Workflow Commission (IWC). While ABRomics currently supports genomic analyses of isolated bacterial strains, its capabilities are now being extended to metagenomics, bypassing the need for prior bacterial isolation. To this end, we have developed three complementary workflows, each providing a distinct level of resolution from metagenomic samples.

Workflow WF1 operates directly on raw sequencing reads to deliver a rapid initial characterization of both resistome and taxonomic composition of a sample. This read-based approach is three to five times faster than the assembly-based alternatives, making it well suited for high-throughput screening.

Workflow WF2 performs gene-level analyses by assembling raw reads into contigs, enabling a more refined exploration of the resistome, taxonomic diversity, functional potential, and mobile genetic elements (MGEs). This additional assembly step comes at a higher computational cost but yields substantially richer biological insights.

Finally, workflow WF3, developed in the context of ELIXIR FAIRyMAGs initiative, reconstructs Metagenome-Assembled Genomes (MAG) and provides a genome-resolved view of the microbial community. MAGs are fully annotated for taxonomy, resistome, mobilome, and functional potential, offering the most comprehensive characterization of ARG genomic context. Although WF2 and WF3 require greater computational time due to their assembly and binning steps, they enable a deeper understanding of the genomic and ecological context of ARGs.

These three workflows were developed within the Galaxy framework by integrating tools for metagenomics and antibiotic resistance (Sylph, DeepARG, Groot, MMseqs2), and they have been validated by the Intergalactic Workflow Commission (IWC). They complement each other in their

analysis results, aiming to extend ABRomics' functionality beyond genomic surveillance. Fully integrated into the ABRomics platform, they provide a comprehensive framework for the metagenomic surveillance of antimicrobial resistance.

Metagenomic raw reads analysis:

<https://iwc.galaxyproject.org/workflow/metagenomic-raw-reads-amr-analysis-main/>

Metagenomic genes catalog:

<https://iwc.galaxyproject.org/workflow/metagenomic-genes-catalogue-main/>

MAGs reconstruction: <https://iwc.galaxyproject.org/workflow/mags-building-main/>

ABRomics-analysis is accessible at <https://analysis.abromics.fr/>

SPOC: Galaxy Single-cell & sPatial Omics Community of practice

VIDEM Pavankumar, SPOC

Poster only

Single-cell and spatial omics technologies have revolutionized biological research, yet their rapid development has created severe computational bottlenecks, including fragmented tools and incompatible data formats. The computational complexity and steep learning curve associated with different single-cell analysis environments often pose a significant hurdle for experimental biologists, making a user-friendly platform like Galaxy essential for data analysis. To address these challenges, the Galaxy Single-cell & sPatial Omics Community of Practice (SPOC) was established in 2022 as a global, community-driven network of users, developers, and trainers collaborating to advance Galaxy's single-cell and spatial omics tools, workflows, and training resources while promoting scalable and FAIR analysis.

A 2025 community update publication highlights the ecosystem's massive growth, now featuring over 175 SPO tools from the Python and R ecosystems, such as Scanpy and Seurat, as well as 120 training resources provided through the Galaxy Training Network (GTN). These resources collectively guide users through the entire data lifecycle, from planning and data collection to processing and analysis, and ultimately to preservation, sharing, and reuse. SPOC has introduced new analysis methods for single-cell and spatial omics, including support for multimodal data integration, image-based and sequence-based spatial transcriptomics approaches, and advanced tasks such as automated cell type annotation, trajectory inference, and cell-cell interaction analysis. The Galaxy SPO tools support a broad range of standard formats used in single-cell and spatial omics. For single-cell data, these include AnnData, MuData for multimodal data, Loom, Seurat for data from the R ecosystem, and SingleCellExperiment from the Bioconductor community. For spatial data, they support structured formats such as SpatialData and Seurat objects along with imaging formats like OME-TIFF. Dedicated conversion tools further allow seamless transformation between these formats, enabling interoperability across Galaxy single-cell tools and workflows, while simplifying data preprocessing and reuse.

SPOC is also closely aligned with European infrastructure efforts through active engagement with the ELIXIR Single-Cell Omics community. As part of the ELIXIR Spatial2Galaxy project, SPOC is developing a scalable, reproducible, and self-contained Galaxy-based platform specifically for spatial transcriptomics. This platform will feature benchmark-validated tools and workflows, along with associated training materials that incorporate best practices. In addition, SPOC contributes to the BioFAIR Pathfinder initiative through a project focused on developing FAIR and AI-ready single-cell workflows and training resources while strengthening interoperability between Galaxy and the Bioconductor ecosystem.

A Galaxy Workflow for Phylogenetic Analysis and Gene–Species Tree Reconciliation in Land Plant Evolution Research

HILTEMANN Saskia, CLARK James W., KENRICK Paul, GRÜNING Björn A., DONOGHUE Philip, VRIES Jan de, RENSING Stefan A.

Poster only

The transition of plant life from aquatic to terrestrial environments marked a pivotal evolutionary milestone that fundamentally transformed Earth’s ecosystems. All existing land plants originate from a common ancestor within the Phragmoplastophyta, a clade that includes land plants and certain streptophyte algae. While the broader outline of this terrestrialization is known, the genomic and morphological traits of the most recent common ancestor (MRCA) of land plants remain unsolved. Here, we intend to investigate the evolution of traits across streptophyte algae and land plants through comparative genomic and phylogenomic analyses.

To support these analyses, we developed a reproducible workflow in the Galaxy platform for phylogenetic and comparative genomic studies. The workflow combines several important analysis steps, including multiple sequence alignment, phylogenetic tree construction, and gene-species tree reconciliation. It allows users to move from gene sequences to evolutionary interpretation simply and transparently.

The workflow is applied using protein sequence data from MAdLandDB (also integrated into Galaxy EU), a curated database of plant and algal species proteomes developed and maintained by us. By integrating different tools into a single Galaxy workflow, we make complex phylogenomic analyses easier to reproduce, reuse, and share with the community.

This poster presents the overall biological question behind the project, the structure of the Galaxy workflow, and how workflow-based analyses can help in studying the evolution of important plant traits such as drought tolerance, stomatal development, and vascular tissues.

First-time Galaxy tool integration: feedback about challenges faced, from two trainees with different academic backgrounds

**KA Abdou Lahat, RUIN Quentin, CLADIERE Mathieu, WORKFLOW4METABOLOMICS
Coreteam, PETERA Mélanie**

Poster only

Integrating existing command-line bioinformatics tools into Galaxy is an important step to make analyses more accessible and reproducible. However, for first-time developers, this process can involve several challenges. In this poster we present two Galaxy tool integrations by two internship students with different academic backgrounds (bioinformatics and applied mathematics). The tools rely on existing metabolomics' algorithms developed by collaborating laboratories and used internally.

We approached the same process in slightly different ways, but both included wrapper design, XML configuration, testing, documentation and addition of new features to improve usability within metabolomic workflows. Across both case studies, similar recurring challenges emerged despite differences in tools and approaches (missing values imputation for Tool 1, and table comparison for Tool 2). For example, these included input/output formatting issues, dependency-related behaviour, adapting scripts not initially designed for Galaxy, choosing clear user parameters and balancing code optimization while preserving the original validated method.

To address these challenges that one may face, we propose feedback about diverse aspects, e.g. the use of virtual environments for dependency control, systematic testing with representative datasets. Based on these observations, we suggest simple strategies to make Galaxy integration smoother. Despite the challenges, both tools were successfully handled, enhanced and tested in order to be made available for use in Galaxy and published on GitHub.

Our experience suggests that Galaxy tool development is accessible to new contributors, but requires special care about supervision, systematic testing and a clear understanding of the tool behaviour. First-time integration experiences can provide useful feedback for improving onboarding practices, in particular in laboratories with no prior experience in Galaxy tool wrapping.

Novel Integrated Galaxy Workflows for comprehensive sncRNA Analysis

UEDA Joe, TEYSSET Laure, CARRÉ Clément, NAOUAR Naïra

Poster only

Small non-coding RNAs (sncRNAs) are a class of regulatory RNAs that play a central role in gene expression and genome stability. Although this field has recently gained significant attention, many aspects of sncRNA biology remain poorly understood, especially regarding their emergence and functional interactions.

As a representative use case, PIWI-interacting RNAs (piRNAs) illustrate the challenges of sncRNA analysis. Although their role in transposable element silencing is well characterized, the processes governing the emergence and activation of piRNA-producing genomic regions, known as piRNA clusters, remain less understood. Their activation can be assessed by quantifying their abundance using small RNA-seq across different conditions, complemented by the analysis of epigenetic markers, such as histone modifications, through ChIP-seq experiments. While several individual analysis tools such as FastQC for quality control, Bowtie for sequence alignment, and MACS3 for peak-calling are already available within the Galaxy platform, important components are still missing, including visualization tools such as Enhanced heatmap. More importantly, dedicated workflows tailored to sncRNA have not been developed yet.

This limitation applies to other sncRNA classes such as tRNA-derived fragments (tRFs), whose elevated levels have been shown to disrupt other sncRNA-mediated gene repression pathways, including piRNAs. However, no specialized tool for tRF quantification, like MINTmap, is available within Galaxy, further limiting accessibility for non-computational users.

In this context, we present two bioinformatics tool suites and associated workflows designed to investigate different aspects of sncRNA biology. One workflow is dedicated to the analysis of ChIP-seq data for the profiling of histone modifications and the second one focuses on the characterization of sncRNA population through small RNA-seq data. They will integrate existing tools, introduce missing functionalities, and improve accessibility through the Galaxy framework.

Small Scale Admins

HOTZ Hans-Rudolf, LOPEZ-DELISLE Lucille, GOUÉ Nadia, BERNT Matthias, ČECH Martin,
DAVIS John, VISAN Vlad

Poster only

The Small Scale Admins group is dedicated to maintainers of local Galaxy servers. Local Galaxy deployments are widely used to support local scientific analysis workflows, yet they require sustained administrative effort to remain functional, scalable, and user-friendly. Since 2021, our Small Scale Admins community (40+ members across 30+ universities/institutes) brings together small scale Galaxy administrators, giving them a forum to exchange practical experience, best practices and address recurring operational challenges across institutions.

Navigating Galaxy's Growth: How CoDex Lower Barriers to Access

CORREARD Solenne, BRETAUDEAU Anthony, ZIEREP Paul, BATUT Bérénice

Poster only

The Galaxy Project offers a vast and growing ecosystem of tools, workflows, and tutorials, empowering researchers to perform accessible, reproducible, and transparent data analysis. However, the sheer volume of resources can overwhelm both new and experienced users, hindering accessibility and adoption.

This poster introduces the Galaxy Communities Dock (CoDex) as a pivotal solution that bridges massive resources and communities. CoDex is a centralized, manually curated repository that aggregates community-specific tools, workflows, and tutorials. Resources are automatically updated weekly. By combining technical infrastructure with human curation, CoDex supports communities in automatically highlighting their most relevant resources, embodying both a people-centered and technical approach. The CoDex is also a key element for the Labs, user-friendly interfaces built on the Galaxy framework, which enable communities to rapidly identify and launch relevant resources on their favorite Galaxy instance without requiring programming expertise.

The poster will highlight the current state of CoDex and Labs, and how it fits in the Galaxy governance and infrastructure, with examples such as the microGalaxy Lab or the Biodiversity-genomics Lab. Additionally, it will outline ongoing and future work, particularly in the context of upcoming CoFest discussions, where the community will focus on refining and expanding these initiatives. To foster collaboration, the poster will also clearly highlight our need for contributors to the CoDex for code review or implementation of new features. This approach aims to engage the community in shaping the future of CoDex and Galaxy's accessibility.

Repository-Coupled scRNA-seq Analysis with Galaxy

OLHA Jaroslav, VALKOVSKÝ Viktor, JURÁČEK Jaroslav, KŘENEK Aleš

Poster only

We present a single-cell RNA-seq (scRNA-seq) data processing pipeline which integrates multiple analysis platforms to support reproducibility and FAIR data management. In collaboration with a medical biology lab, we developed the pipeline to address practical barriers that prevent researchers from producing reusable datasets and analysis workflows.

The pipeline connects multiple storage and analysis platforms – raw data are first retrieved from Illumina BaseSpace and processed by Parse Biosciences Trailmaker in order to generate count matrices and the corresponding AnnData objects. These are then processed by a predefined Galaxy workflow. All the important intermediate data and final outputs – including all parameter settings, quality control reports, filtered and unfiltered count matrices and final visualizations – are stored in an Invenio-based repository maintained as part of the Czech National Repository Platform.

All data, analysis workflows and results are bundled as RO-Crate packages, enabling full reproducibility of published results from raw inputs, and reusability of workflows for later analyses. The pipeline demonstrates how direct integration of Galaxy workflows with repository infrastructure can automate provenance capture and long-term archival of scRNA-seq analyses.

The Galaxy community in Switzerland

LOPEZ-DELISLE Lucille, HOTZ Hans-Rudolf, THANKAM SREEDHARAN Vipin

Poster only

Only few people have access to a local (aka Swiss based) galaxy server in Switzerland. Nevertheless, the usage of usegalaxy.eu by people registering with a “*.ch” mail address is huge (more than 1000 registrations as of May 2026). The aim of this community is to make things easier for any people in Switzerland wishing to contact other Swiss people.

Processing zebrafish (*Danio rerio*) video and image data using Galaxy: A web based solution for toxicology and ecotoxicology research

MASSEI Riccardo, BERNT Matthias, NICOLAY Elena Katharina

Poster only

Zebrafish (*Danio rerio*) embryos are among the most commonly used alternative animal models for fundamental research and for biomedical applications. In particular, the use of zebrafish in toxicology and ecotoxicology can greatly help to reduce animal testing and endorse the use of alternative approaches for chemical testing. Due to the relative transparency of the larval body, several microscopy techniques can be used to study its anatomical structures, organ development but also development of basic behavioral traits. To this end, several image and video analysis techniques have been developed in the recent years to study the effect of chemical exposure and environmental pollution on swimming behavior and organ development such as eyes and body length. These approaches are normally based on processing image and video data using in-house software or relying on existing Python or R libraries. Such approaches can be limiting for users who are not used to programming or not having access to proprietary software or to required compute infrastructure for data analysis. In this work, we are showcasing how to perform a series of video and image analyses using Galaxy as open web-platform for reproducible science together with existing video and imaging processing tools. This includes body and organ segmentation using AI approaches (i.e. Cellpose), video frame extraction, analysis of deep-learning (DL) model training, run Python code in Jupyter Notebooks or perform simple dose-response curves modeling. Thanks to the Galaxy framework, reproducible workflows in zebrafish research can be re-used by other scientists worldwide, promoting science reproducibility and transparency.

Galaxy-BioProd: An Integrated FAIR Platform for Synthetic Biology, Biotechnology and Life Cycle Analysis

PRAGASSAM Anthony, CHAUSSEPIED Thomas, BRETAUDEAU Anthony, LE CORGUILLÉ Gildas, FAULON Jean-Loup, LOUX Valentin

Poster only

The "Bioproductions (B-BEST): Biomass, Biotechnologies, and Sustainable Technologies for Chemistry and Fuels" research program is part of the national acceleration strategy "Bio-based Products and Industrial Biotechnologies - Sustainable Fuels" under the France 2030 plan. Co-led by INRAE and IFPEN, this program supports the transition from a petrochemical economy to bio-based products and services by aiming to better understand and utilise biomass to produce bio-based products and sustainable fuels.

A secondary objective is to establish common practices for bio-refineries, which are currently not implemented at national level, thereby limiting the efficiency and reproducibility of developments. To address these issues, the Galaxy-BioProd project aims to develop a centralised portal providing standardised digital tools and resources to simplify, connect and accelerate industrial biotechnology projects. The tools and pipelines are developed by the members of the project.

Based on the Galaxy environment and adhering to FAIR principles, the B-BEST Galaxy lab deployed on usegalaxy.fr offers an accessible and unified platform that can serve various communities in the fields of synthetic biology, biocatalysis and industrial biotechnology.

Seventeen tools ranging from synthetic biology to Life Cycle Analysis are currently under development and will be deployed on B-BEST lab of usegalaxy.fr.

Galaxy-BioProd project is part of the PEPR Bioproductions (B-BEST) program, supported by the France 2030 investment plan and co-led by INRAE and IFPEN [ANR-22-PEBB-0008].

GaaS: Galaxy as a Service

PLANEL Rémi

Poster only

Background

The Galaxy Project is a cornerstone of bioinformatics, providing a comprehensive suite of tools for data-intensive analysis. While its REST API enables powerful automation, the standard web interface can overwhelm end users who need to execute specific workflows. Although the API supports custom web interfaces, building them presents significant practical challenges: interacting with Galaxy's rich API surface, implementing authentication and permissions, and managing persistent data storage.

Results

We present GaaS (Galaxy as a Service), a TypeScript toolkit (currently in beta) designed to streamline the creation of custom Galaxy-powered web applications. GaaS leverages the Nuxt/Vue ecosystem to handle infrastructure complexity, enabling developers to focus on domain-specific visualizations and user experiences. The toolkit uses Galaxy to reproducibly describe, run, and manage workflows through a structured, programmatic interface.

The GaaS ecosystem comprises four integrated packages:

- **blendtype** — A TypeScript client built with Effect-TS that provides a robust, type-safe interface for Galaxy's REST API.
- **nuxt-galaxy** — A Nuxt module that abstracts Galaxy integration, including file storage, user authentication, permissions, and database schema via Supabase.
- **gaas-ui** — A Nuxt layer providing composable, intuitive web interface components tailored for bioinformatics workflows.
- **wiki** — A content layer providing domain-specific documentation that helps users interpret their analysis results and understand the scientific context of their workflows.

GaaS is currently powering our internal satelliteFinder webservice, demonstrating its viability for production bioinformatics applications.

Conclusion

GaaS significantly reduces the barrier to building dedicated web interfaces for Galaxy tools and workflows. By abstracting API complexity, authentication, and data persistence with a reproducible workflow engine, it enables developers to concentrate on building rich, interactive visualization and exploration interfaces for research data, making sophisticated computational biology accessible to domain experts.

Towards an ODATIS Galaxy Environment for Reproducible Cross-Domain Marine Data Analysis

LEROI Laura, JOSSE Marie, CAER Gwenaël, DETOC Jérôme

Poster only

Towards an ODATIS Galaxy Environment for Reproducible Cross-Domain Marine Data Analysis

As a national ocean data and services hub, ODATIS is exploring how Galaxy could be integrated into its service catalog as an open solution for reproducible marine data analysis, workflow sharing, and training activities. This initiative builds upon the work carried out within the FAIR-EASE project in collaboration with the Galaxy Europe team, including the development of Galaxy for Earth System Science as a key outcome, which contributed to extending Galaxy capabilities toward marine and environmental sciences. This work explores a Galaxy-based environment for reproducible marine data analysis, integrating multi-source data, scalable workflows, and High Performance Computing (HPC) access at the institutional level.

Within FAIR-EASE and the Galaxy for Earth System Science initiative, several new functionalities were developed to support ocean-related scientific workflows, including simplified access to oceanographic and satellite datasets, integration of interactive visualization environments, harmonization of heterogeneous observations, spatial interpolation and mapping of ocean variables, estimation of biogeochemical parameters using machine learning, detection of biosynthetic gene clusters from marine omics data, and the extension of workflows toward AI-based analysis, including YOLO-based models applied to marine imagery for species detection.

ODATIS aims to capitalize on these developments and continue deploying operational Galaxy-based services for reproducible and accessible marine data analysis, facilitating the transition from isolated notebook-based developments toward reusable, reproducible, and scalable Galaxy workflows accessible to scientists with limited IT expertise, while enabling the integration and cross-analysis of oceanographic, satellite, biological, omics, and bioinformatics data within shared analytical pipelines. This cross-domain capability is particularly relevant for marine science, where understanding ecosystem functioning increasingly requires connecting physical and biogeochemical observations with biodiversity, microbiome, genomic, and functional data. In this context, ODATIS also aims to simplify access to the HPC infrastructure of Ifremer's Marine Data and Computing Center through Galaxy-based interfaces and workflows, enabling scalable processing of large marine datasets.

This initiative continues FAIR-EASE and Galaxy for Earth System Science efforts, while contributing to the Galaxy community through shared tools, workflows, and training materials via the Galaxy Training Network (GTN), and by promoting interoperability with larger-scale platforms and distributed infrastructures, notably through Pulsar.

The poster will present several representative applications, including:

- satellite and in situ data colocalization workflows;
- coastal and ocean biogeochemical data analysis workflows;
- marine biodiversity and metabarcoding analysis workflows;
- marine metagenomics and microbiome analysis workflows;
- integration with the HPC infrastructure of Ifremer's Marine Data and Computing Center;
- as well as an overview of marine science tools integrated into the Galaxy ecosystem.

Three Galaxies, One Grid, One Foundation: Building National Bioinformatics Infrastructure in Czechia

CECH Martin, DEMKO Martin

Poster only

In Czechia, CESNET and ELIXIR Czechia jointly operate three independent Galaxy instances serving distinct scientific communities: RepeatExplorer for genomics researchers working with repetitive elements, UMSA for the RECETOX community with untargeted mass spectrometry and cheminformatics pipelines, and usegalaxy.cz as a general-purpose instance bridging users to MetaCentrum, the Czech national computational grid. Despite their different audiences, all three are deployed from a unified, publicly available infrastructure cookbook, ensuring consistent configuration across heterogeneous environments. Together, the three instances have accumulated over 1,800 registered users and in 2025 executed 115,000 jobs consuming 108,000 CPU-days, demonstrating sustained production use at national scale.

Access across all three instances is federated through national e-infraCZ Login and Life Science Login. Tool updates are managed via an automated, CI-validated pipeline deployed with Ansible, with toolsets maintained in a public repository - an approach adapted from practices developed by other Galaxy administrators and generalized into a reusable template to streamline tool management for other instances, publicly available at https://github.com/martenson/galaxy_tools_management_template.

A recurring operational challenge has been connecting Galaxy to MetaCentrum's PBSpro job scheduler via Pulsar; we will share the custom solutions developed and the limitations that remain, including the lack of job rescheduling support - lessons relevant to any deployment targeting national HPC infrastructure.

We are currently piloting integration of Galaxy with the National Repository Platform (NRP), a CESNET-led initiative built on InvenioRDM, the same technology underlying Zenodo, that provides the core research data infrastructure for Czechia. Galaxy instances are being connected to NRP repositories as native file sources, enabling researchers to pull datasets into analyses and deposit results directly from within their analytical environment. We will report on early findings from this pilot and discuss the potential for this approach to connect Galaxy-based computation with national FAIR data infrastructure more broadly.

Launching the Galaxy UK Special Interest Group: National ecosystem coordination supported by the BioFAIR Fellowship

SORANZO Nicola, LOACH Marisa

Poster only

The UK bioscience community features a highly active and diverse ecosystem of Galaxy users, tool and workflow developers, trainers, and service providers. To coordinate these efforts, we are establishing a Galaxy UK Special Interest Group (SIG) under the official Galaxy community governance, with support from the BioFAIR Fellowship programme. BioFAIR is a £34 million UKRI-funded digital research infrastructure designed to connect existing services, communities, and expertise into a cohesive national landscape for FAIR (Findable, Accessible, Interoperable, Reusable) data management and analysis for the UK life sciences. Within this framework, the BioFAIR Fellowship provides dedicated specialist time to engage with the UK-wide community and build a national "People Commons", with the first cohort focused on workflow development and management. Here we present how this collaborative approach is driving community-led governance and promoting practical workflow FAIRification both domestically and internationally.

We began by engaging researchers and technical staff at key UK research organisations actively involved in Galaxy use, development, or training. Convened for its inaugural meeting in May 2026, the SIG has successfully brought regional stakeholders together to establish the group's web presence, launch a public Matrix chat channel, and discuss formal representation on the global Galaxy Community Board. The SIG serves as an open, welcoming forum to discuss national training and infrastructure needs, helping to shape the services being built by BioFAIR, in particular its Methods Commons. This integrated digital platform is designed to provide life scientists with access to national-scale capabilities for the discovery, execution, sharing and reuse of computational workflows, tools and notebooks.

Alongside national governance and requirements gathering, the Fellows support the community in adopting reproducible, open science practices by assisting researchers in aligning their Galaxy workflows with FAIR standards and publishing them via registries like the Intergalactic Workflow Commission (IWC) and WorkflowHub. These dissemination efforts have been promoted internationally, such as through the talk "FAIR computational workflows and Galaxy" delivered at the March 2026 online meeting of the Euro-BioImaging FAIR Image Data Workflows Expert Group, and contributing to the online Galaxy Training Academy (18-22 May 2026). Domestically, these practical concepts are being advanced through dedicated community events, including a breakout session at the BioFAIR Showcase in Cambridge (14-15 May 2026). This continuous engagement through the SIG and practical workshops ensures that the development of BioFAIR's national infrastructure remains rooted in the real-world requirements of the UK research community.

SemSec: Using Crypt4GH Recryption to Secure and Orchestrate Analysis of Encrypted Datasets from EGA/FEGA in Galaxy

GUNDERSEN Sveinung, CHAVERO DÍEZ María, DE GEEST Paul, DE LA TORRE PERNAS Sabela, MAIER Wolfgang, REPCHEVSKY Dmitry, MOKHTAR FARAG Yehia, LÓPEZ TABERNERO David, NIEDERLÖHNER Ralph, VÁZQUEZ Pável, AASLAND Rein, COPPENS Frederik, GRÜNING Björn, HOVIG Eivind, RAMBLA Jordi, CAPELLA-GUTIERREZ Salvador

Poster only

The growing use of controlled-access human data and encryption-at-rest policies in biomedical research creates tension between strong security and interactive, workflow-based analysis. Virtual Research Environments (VREs) such as Galaxy are widely adopted for reproducible, user-friendly data analysis, but their traditional execution model assumes that data can be decrypted within the platform or its attached compute resources. For sensitive data, this assumption is no longer acceptable: private keys must remain under user control, and decryption should be confined to Trusted or Secure Processing Environments (TREs/SPEs).

Within the ELIXIR Commissioned Service “Empowering Users: Orchestrating Sensitive Data Access for Interactive Federated Analysis in VREs”, we are developing a solution to allow Galaxy to securely analyse datasets encrypted according to the GA4GH Crypt4GH standard. At its core is a Crypt4GH-based Recryptor architecture that allows Galaxy to orchestrate analysis of encrypted data without ever handling user private keys or plaintext. With direct support for importing encrypted datasets from the European Genome-phenome Archive (EGA, including Federated EGA), EGA/FEGA serves as the primary use case. Our approach is, however, generalisable not only to other sources of encrypted data, but also to other VREs.

We treat encrypted files as first-class objects in Galaxy via Crypt4GH-aware datatypes that maintain the specificity needed to link tools and workflows. We take advantage of a powerful feature of Crypt4GH to re-encrypt (or “recrypt”) small headers without touching the encrypted data itself. In Galaxy, we store recrypted headers in the metadata of the datasets, allowing the dataset to remain untouched wherever it is located. A user-side Recryptor service runs locally, exposing a REST API that the Galaxy web client contacts from the browser. This service stores the user’s Crypt4GH private key locally and performs header-only “recryption” to short-lived public keys of compute-side Recryptor services deployed in TRE/SPE backends. Galaxy’s existing remote file source and job scheduling mechanisms then stream encrypted payloads directly to secure compute nodes, where decryption, analysis, and re-encryption of outputs for the authorised user are performed.

By GCC 2026, we plan to demonstrate this pattern end-to-end on synthetic, human-data-like datasets, including an EGA integration, subject to the progress of ongoing development. The result is a reusable design that preserves Galaxy’s interactive strengths while aligning with modern requirements for encrypted, federated analysis of sensitive data.

ProMaya: a hierarchical universal Deep Learning framework for accurate and interpretable Protein-Protein interaction identification

BHATI Umesh, SHANKAR Ravi

Poster only

Protein-protein interactions (PPIs) are molecular lego which define the physical states of cells. Accurately identifying PPIs remains challenging due to the interplay of several factors ranging from electrostatic to molecular geometry, topology, and physics. Existing computational approaches capture only fragments of this orchestra, limiting their generalizability across protein families and interaction types. Here, we present ProMaya, a hierarchical multi-scale Graph-transformer framework which has learned the 3D atomic geometry, electronic distribution, residue-level structure and disorder, surface mass-density signatures, and large protein language-model embeddings of interacting proteins. Highly comprehensively benchmarked across nine species and 47 GB experimentally validated data, ProMaya achieved consistently >95% average accuracy, outperforming state-of-the-art tools by >12%. Its in-built explainability provides clear mechanistic reasoning for the observed interactions. The first time introduced atomic and protein language learning has dramatically made it attain an outstanding level for PPI discovery in absolutely universal manner which can deal any species and potent to even bypass costly experiments. ProMaya system is freely accessible at <https://scbb.ihbt.res.in/ProMaya/>

Reproducible and Secure Galaxy environments for the EOSC community

MINKOVA Mirela

Demo only

Open and FAIR Science is a core research priority, yet its implementation becomes challenging when working with General Data Protection Regulation (GDPR) data. Within the European Open Science Cloud (EOSC) ecosystem, Dutch infrastructures such as EOSC NL and SURF Research Cloud (SRC) provide a way to analyse sensitive data in secure, on-demand environments while still adhering to FAIR practices.

Through EOSC NL, researchers can launch reproducible Galaxy environments for secure analysis without the burden of manual installation, maintenance, or updates. The same deployment can be reused across projects, providing teams with a consistent and accessible infrastructure while still allowing flexibility in choices such as Galaxy version and environment size to meet each project's specific needs. Access is managed through secure authentication linked to project rights, and administrator coupling provides the governance required in restricted settings. Clear boundaries are established around deployment, access, and instance management.

A key strength of this setup is its portability within the EOSC NL ecosystem. Galaxy and Pulsar can be deployed as separate components across different cloud providers inside EOSC NL, allowing Galaxy to be positioned where user access and data handling requirements are best met, while Pulsar runs where compute capacity or policy constraints demand it. This separation enables scaling across multiple virtual machines and providers while keeping the analysis experience consistent for researchers. Flexible deployment profiles allow environments to be right-sized for the workload, including CPU and RAM configurations, and GPU support where relevant.

The end-to-end workflow starts by launching a Galaxy instance through EOSC NL services, selecting the version and resource profile, connecting to Pulsar with minimal configuration, and running an analysis in secure, GDPR compliant environments.

SemSec: Using Crypt4GH Recryption to Secure and Orchestrate Analysis of Encrypted Datasets from EGA/FEGA in Galaxy

GUNDERSEN Sveinung, THE SEMSEC TEAM

Demo only

The growing use of controlled-access human data and encryption-at-rest policies in biomedical research creates tension between strong security and interactive, workflow-based analysis. Virtual Research Environments (VREs) such as Galaxy are widely adopted for reproducible, user-friendly data analysis, but their traditional execution model assumes that data can be decrypted within the platform or its attached compute resources. For sensitive data, this assumption is no longer acceptable: private keys must remain under user control, and decryption should be confined to Trusted or Secure Processing Environments (TREs/SPEs).

Within the ELIXIR Commissioned Service “Empowering Users: Orchestrating Sensitive Data Access for Interactive Federated Analysis in VREs”, we are developing a solution to allow Galaxy to securely analyse datasets encrypted according to the GA4GH Crypt4GH standard. At its core is a Crypt4GH-based Recryptor architecture that allows Galaxy to orchestrate analysis of encrypted data without ever handling user private keys or plaintext. With direct support for importing encrypted datasets from the European Genome-phenome Archive (EGA, including Federated EGA), EGA/FEGA serves as the primary use case. Our approach is, however, generalisable not only to other sources of encrypted data, but also to other VREs.

We treat encrypted files as first-class objects in Galaxy via Crypt4GH-aware datatypes that maintain the specificity needed to link tools and workflows. We take advantage of a powerful feature of Crypt4GH to re-encrypt (or “recrypt”) small headers without touching the encrypted data itself. In Galaxy, we store recrypted headers in the metadata of the datasets, allowing the dataset to remain untouched wherever it is located. A user-side Recryptor service runs locally, exposing a REST API that the Galaxy web client contacts from the browser. This service stores the user’s Crypt4GH private key locally and performs header-only “recryption” to short-lived public keys of compute-side Recryptor services deployed in TRE/SPE backends. Galaxy’s existing remote file source and job scheduling mechanisms then stream encrypted payloads directly to secure compute nodes, where decryption, analysis, and re-encryption of outputs for the authorised user are performed.

By GCC 2026, we plan to demonstrate this pattern end-to-end on synthetic, human-data-like datasets, including an EGA integration, subject to the progress of ongoing development. The result is a reusable design that preserves Galaxy’s interactive strengths while aligning with modern requirements for encrypted, federated analysis of sensitive data.