Keywords:
#Interdisciplinary, #OpenScience, #Providers, #HPC, #OpenScienceLabs, #ResearchCommunity, #EOSCinPractice

Empowering Research Communities through cross-sector data sharing

Enabling Collaboration and Advancements in High-Performance Compute through Open Science Lab

Constanze Roedig

Head of the Austrian DataLAB and the Austrian Open Cloud Community, Vienna Scientific Cluster

"By lowering the entrance barrier, our vision is to make High Performance Computing accessible, usable and approachable by scientists across disciplines, students and the general public. Sharing insights across fields is necessary for truly advancing rather than constantly reinventing the wheel."

OCRE & EOSC

The Open Clouds for Research Environments project (OCRE), aims to accelerate cloud adoption in the European research community, by bringing together cloud providers, Earth Observation (EO) organisations and the research and education community. OCRE will make selected commercial digital services an integral part of the European Open Science Cloud (EOSC), ensuring compliance with EOSC requirements and visibility in the EOSC-hub Service Catalogue.

The service provider

The Vienna Scientific Cluster (VSC) provides High-Performance Compute services to all of Austria and acts as a PRACE training centre. To further enhance Research and Research-Driven Teaching, they have begun incubating Cloud Services on top of High-Performance Compute Infrastructures. The Austrian Datalab and Services Project is developing an Austrian wide federated cloud-infrastructure via using automation and DevSecOps, and is maintaining infrastructure blueprints. Sparkle and Technische Universität Wien are using EOSC Future INFRAEOSC-03-2020 Grant Agreement Number 101017536 to make these services available on EOSC.

The user Community

Research groups in traditional High Performance Compute fields such as Physics, Chemistry, Geoscience can use this service to host these interactive Labs for Outreach, Summer schools and Onboarding. Also emerging communities like Machine Learning and DataScience can benefit greatly by using the Analysis-Labs. On the development side, they are also an active research topic for CyberSecurity and OpenScience design principles.

The user Perspective

Master and PhD students can use real HPC results data together with the original tooling (e.g., bespoke coordinate transformation libraries) to experiment with scientific questions or conduct data analysis without being exposed to technical problems like compilers or dependency management.

Why do I need EOSC?

EOSC provides the essential central location to discover, publish and consume the HPC labs and results, which is key for any opensource and/or FAIR project. Furthermore, our intent is the standardisation or, at least, collaborative co-development of common low-level infrastructure services, and for this we require access to EOSC-core like AAI but also an integration with Helpdesk.

The Challenge

The sharing of insights in science and the passing on of knowledge is an ever-increasing issue due to the proliferation of scientific subdomains and the growing specialisation of each sub-field. For scientists themselves, the effort required to stay up-to-date within their own field is growing. Sharing insights across fields is necessary for truly advancing rather than constantly reinventing the wheel. Thus, we require methods aiding this translation and reducing the toil on the scientist, to share, teach, collaborate, peer-review, learn and publish.
Our vision is to use commoditise access to science and to encourage newcomers by lowering the learning curve and cognitive friction in the field of High-Performance Computing.

The EOSC Future project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020 - Grant Agreement Number 101017536

EOSC Future

OCRE

EOSC in practice story #18

Keywords:
#Interdisciplinary, #OpenScience, #Providers, #HPC, #OpenScienceLabs, #ResearchCommunity, #EOSCinPractice

The Solution

OpenScienceLabs (OSL) enable the sharing of HPC code and results using cloud technology. It provides a universal wrapper for running code in the browser, with access to parallel or special purpose hardware. OSL utilises cloud-native methods and DevSecOps practices, such as containerisation, orchestration, automation, and security measures. It simplifies usage for new or non-expert users by eliminating the need for compilation, installation, or operational tasks.
With OSL, end-users can access the system within minutes through a web browser. After an initial co-development phase, any HPC codebase can be made compatible with OSL. The OSL design is continuously improved through use-cases like Peer Review-Companions, Outreach, Open Sourcing, Self-Service On-boarding, and HPC-Summer Schools. These use-cases help enhance usability, scalability, and portability of OSL.

Useful tips & tricks

The "OpenScienceLabs for HPC" are based on open-source blueprints available here: https://dev.azure.com/AOCC/OpenScienceLabs
We implemented an opinionated, hardened, but lightweight Kubernetes stack combined with a generic wrapping mechanism (using JupyterHub) and modern Identity Federation. Our approach focuses on separating state and run-time, enhancing security and enabling independent development of Labs and user data.

The impact on society

Our vision is to democratise High Performance Computing by making it more accessible and usable for scientists, students, and the general public. We aim to lower the entrance barrier, relieve researchers from technical burdens, simplify peer-review processes, and establish a digitally sovereign infrastructure in academic institutions. Through "Open Research Platforms in the Cloud," universities can collaborate using our secure, affordable, and tested OpenSource blueprints. This collaborative effort will improve cost-efficiency, save researchers time, democratise HPC knowledge, and encourage serendipitous discoveries.

Useful material related to this story

ocre-project.eu

open-science-labs

Across disciplines

The OpenScienceLabs for HPC is most obviously encouraging the cross-disciplinary research between data science, machine learning and STEM subjects and was built on that premise. Furthermore, research in computer science and embedded engineering will likely be combined in order to improve the security and accessibility.

Sustainability for an EOSC in practice

Most Open Science Labs for HPC are initially development-heavy for a Research Group. However, once established, we plan on offering workshops and blueprints to maintain the operations in their own cloud-native infrastructures. This may require establishing those at various HPC facilities. We are conducting active research on parametrised operational models to quickly estimate the financial feasibility of a use case.

Future developments & funding scenarios

After completing the 3D rendering features, we plan to integrate Machine Learning Workflow tools like Kubeflow/Elyra. This will enable connecting workflow steps to HPC clusters, enhancing security measures, and improving the self-service aspect of our offering. We also aim to provide more blueprints and initial training to Research Groups, although we are currently in the prototype stage.
For future funding models, we are considering shifting OpenScienceLabs more towards teaching. This would reduce operational costs and cloud spending while maintaining the production of open-source artifacts. It could also inspire students to pursue R&D careers. Additionally, we are exploring the possibility of a run-time federation with existing HPC infrastructures, although this poses technical challenges.

Liked this #EOSCinPractice story?

Follow @EOSCFuture for more!

Share your own #EOSCinPractice story here