Contribution ID: **104**   Type: **Short Talk**

Collaborative research in the cloud with Notebook-as-a-VRE (NaaVRE)

Wednesday, 2 October 2024 17:15 (10 minutes)

Many scientific problems, such as environmental research or cancer diagnosis, require large data volumes, advanced statistical or AI models, and distributed computing resources.

To help domain scientists conduct their research more effectively they need to reuse resources like data, AI models, workflows, and services from different sources to address complex challenges. Sharing resources requires collaborative platforms that facilitate advanced data science research that offers: discovery access, interoperation and reuse of research assets, and integration of all resources into cohesive observational, experimental, and simulation investigations with replicable workflows. Virtual Research Environments (VREs) effectively supported such use cases offering software tools and functional modules for research management. However, while effective for specific scientific communities, existing VREs often lack adaptability and require substantial time investment for incorporating external resources or custom tools. In contrast, many researchers and data scientists prefer notebook environments like Jupyter for their flexibility and familiarity.

To bridge this gap we propose a VRE solution for Jupyter Notebook-as-a-VRE (NaaVRE).

The NaaVRE empowers users to construct functional blocks by containerizing cells within notebooks, organizing them into workflows, and overseeing the entire experiment cycle along with its generated data. These functional blocks, workflows, and data can then be shared within a common marketplace, fostering user communities and tailored Virtual Research Environments (VREs). Additionally, NaaVRE seamlessly integrates with external repositories, enabling users to explore, select, and reuse various assets such as data, software, and algorithms. Lastly, NaaVRE is designed to seamlessly operate within cloud infrastructures, offering users the flexibility and cost efficiency of utilizing computational resources as needed.

We showcase the versatility of NaaVRE by building several customized VREs that support specific scientific workflows across different communities. These include tasks such as extracting ecosystem structures from Light Detection and Ranging (LiDAR) data, monitoring bird migrations via radar observations, and analyzing phytoplankton species. Additionally, NaaVRE finds application in developing Digital Twins for ecosystems as part of the Dutch NWO LTER-LIFE project.

Topic

Needs and solutions in scientific computing: Platforms and gateway

Primary authors: PELOUZE, Gabriel (LifeWatch ERIC, Virtual Lab & Innovation Center (VLIC), Amsterdam, The Netherlands); KOULOUZIS, Spiros (LifeWatch ERIC, Virtual Lab & Innovation Center (VLIC), Amsterdam, The Netherlands); ZHAO, Zhiming (Multiscale Networked Systems, University of Amsterdam, Amsterdam, The Netherlands)

Presenter: PELOUZE, Gabriel (LifeWatch ERIC, Virtual Lab & Innovation Center (VLIC), Amsterdam, The Netherlands)

Session Classification: Simplifying Data-Driven Science with User-Friendly Platforms and Gateways