

EGU24-17978, updated on 03 Nov 2024

<https://doi.org/10.5194/egusphere-egu24-17978>

EGU General Assembly 2024

© Author(s) 2024. This work is distributed under the Creative Commons Attribution 4.0 License.



Notebook-as-a-VRE (NaaVRE): From private notebooks to a collaborative cloud virtual research environment

Gabriel Pelouze^{1,2}, Spiros Koulouzis^{1,2}, and Zhiming Zhao^{1,2}

¹LifeWatch ERIC, Virtual Lab & Innovation Center (VLIC), Amsterdam, The Netherlands

²Multiscale Networked Systems, University of Amsterdam, Amsterdam, The Netherlands

Studying many scientific problems, such as environmental challenges or cancer diagnosis, requires extensive data, advanced models, and distributed computing resources. Researchers often reuse assets (e.g. data, AI models, workflows, and services) from different parties to tackle these issues. This requires effective collaborative environments that enable advanced data science research: discovery access, interoperation and reuse of research assets, and integration of all resources into cohesive observational, experimental, and simulation investigations with replicable workflows. Such use cases can be effectively supported by Virtual Research Environments (VREs). Existing VRE solutions are often built with preconfigured data sources, software tools, and functional components for managing research activities. While such integrated solutions can effectively serve a specific scientific community, they often lack flexibility and require significant time investment to use external assets, build new tools, or integrate with other services. In contrast, many researchers and data scientists are familiar with notebook environments such as Jupyter.

We propose a VRE solution for Jupyter to bridge this gap: Notebook-as-a-VRE (NaaVRE). At its core, NaaVRE allows users to build functional blocks by containerizing cells of notebooks, composing them into workflows, and managing the lifecycle of experiments and resulting data. The functional blocks, workflows, and resulting datasets can be shared to a common marketplace, enabling the creation of communities of users and customized VREs. Furthermore, NaaVRE integrates with external sources, allowing users to search, select, and reuse assets such as data, software, and algorithms. Finally, NaaVRE natively works with modern cloud technologies, making it possible to use compute resources flexibly and cost-effectively.

We demonstrate the versatility of NaaVRE by building several customized VREs that support legacy scientific workflows from different communities. This includes the derivation of ecosystem structure from Light Detection and Ranging (LiDAR) data, the tracking of bird migrations from radar observations, and the characterization of phytoplankton species. The NaaVRE is also being used to build Digital Twins of ecosystems in the Dutch NWO LTER-LIFE project.