Contribution ID: **124**   Type: **Talk**

Data discovery, analysis and reproducibility in Virtual Research Environments

During the ESCAPE project, the pillars of a pilot analysis facility were built following a bottom-up approach, in collaboration with all the partners of the project. As a result, the CERN Virtual Research Environment (VRE) initiative proposed a workspace that facilitates the access to the data in the ESCAPE Data Lake, a large scale data management system defined by Rucio, along with the interactive analysis that jupyter notebooks enable. The VRE also provisions a variety of scientific software stacks via CVMFS, and can be connected to local data processing resources through REANA. The latter is an open source software, developed within the CERN IT department, that provides a framework focussed on the reanalysis and reproducibility of scientific results. The CERN VRE has deployed an instance of REANA, allowing users to make use of the platform functionalities together with the rest of the services in the analysis facility.

Having a single interface that integrates different services with the underlying infrastructure certainly eases the user experience. Furthermore, in line with the ESCAPE Open Collaboration, the development of open source tools that can be reused in different physics communities with similar analysis strategies would lay the foundation of common lifecycle analysis practices. Therefore, in order to foster accessibility, as well as interactively and reproducibility to more complex infrastructure services, the development of user-friendly middleware should be prioritized.

This contribution focuses on the connection of REANA to the CERN VRE's interface through a Jupyter extension. The development of this extension makes it possible to use the VRE as a single workspace to enhance the lifecycle of a research analysis: from discovery and data access, through interactive analysis and offload to computing resources, to reproducibility of results.

Primary author: GARCIA GARCIA, Enrique (CERN)

Co-authors: GUERRIERI, Giovanni (CERN); GONZALEZ LABRADOR, Hugo (CERN); ESPINAL, Xavier (CERN)

Presenter: GARCIA GARCIA, Enrique (CERN)

Session Classification: Parallel (Track 9)

Track Classification: Track 9 - Analysis facilities and interactive computing