Contribution ID: **550**   Type: **presentation**

REANA: A System for Reusable Research Data Analyses

Monday, 9 July 2018 15:15 (15 minutes)

The revalidation, reinterpretation and reuse of research data analyses requires having access to the original computing environment, the experimental datasets, the analysis software, and the computational workflow steps which were used by the researcher to produce the original scientific results in the first place.

REANA (=Reusable Analyses) is a nascent platform enabling researchers to structure their research data analyses in view of enabling future reuse. The analysis is described by means of a YAML file that captures sufficient information about the analysis assets, parameters and processes. The REANA platform offers a set of micro-services permitting to launch and monitor container-based computational workflow jobs on the cloud. The REANA user interface and the command-line client enables researchers to easily rerun analysis workflows with new input parameters. The REANA platform aims at supporting several container technologies (Docker), workflow engines (CWL, Yadage), shared storage systems (Ceph, EOS) and compute cloud infrastructures (Kubernetes/OpenStack, HTCondor) used by the community.

REANA was developed with the particle physics use case in mind and profits from synergies with general reusable research data analysis patterns in other scientific disciplines, such as bioinformatics and life sciences.

Primary authors: HEINRICH, Lukas Alexander (New York University (US)); Mr HIRVONSALO, Harri (University of Oulu (FI)); RODRIGUEZ RODRIGUEZ, Diego (Universidad de Oviedo (ES)); SIMKO, Tibor (CERN)

Presenter: SIMKO, Tibor (CERN)

Session Classification: T6 - Machine learning and physics analysis

Track Classification: Track 6 −Machine learning and physics analysis