



Contribution ID: 32

Type: **Presentation**

Reproducible high energy physics analyses

Tuesday, January 30, 2018 9:20 AM (20 minutes)

The revalidation, reuse and reinterpretation of data analyses requires having access to the original **virtual environments, datasets, software, instructions** and **workflow steps** which were used by the researcher to produce the original scientific results in the first place. The **CERN Analysis Preservation** pilot project is developing a set of tools that assist the particle physics researchers in structuring their analyses so that **preserving** and **capturing** the knowledge around analyses would lead to easier **sharing, reusing** and **reinterpreting** data. Assuming the full preservation of the original analysis environment, the user code and the computational workflow steps, the **REANA Reusable Analysis** platform enables one to launch **container**-based processes on the computing cloud (Docker, Kubernetes) and to rerun the analysis workflow jobs with new input. The REANA system aims at supporting several **workflow engines** (CWL, Yadage), several **shared storage** systems (Ceph, EOS) and **compute cloud** infrastructures (OpenStack, HTCondor). REANA was developed with the particle physics use case in mind and profits from synergies with general research data analysis patterns in other scientific disciplines such as life sciences.

Primary authors: SIMKO, Tibor (CERN); RODRIGUEZ RODRIGUEZ, Diego (Universidad de Oviedo (ES)); GONZALEZ LOPEZ, Jose Benito (CERN)

Presenters: SIMKO, Tibor (CERN); RODRIGUEZ RODRIGUEZ, Diego (Universidad de Oviedo (ES))

Session Classification: Cloud Infrastructure&Software Stacks for Data Science