Integrating a Lesson on REANA into the HSF Analysis Preservation Training

Gabriel J. Soto
UC Davis (Fellow)
Tibor Simko and Kilian Lieret
CERN (Mentor) and Princeton University (Co-mentor)

Abstract:

The reuse of HEP research data studies requires access to the original analysis environment, datasets, software, and computational workflow. The REANA platform aims to not only store the procedures used to produce original scientific results; but also support container technologies, workflow engines, shared storage systems, and compute cloud infrastructures used by the community. This project will improve the REANA tutorials and documentation to increase the adoption of REANA in the community.

Introduction:

REANA is a widely used HEP program for reproducing data analysis. The software itself allows scientists to run applications in secluded runtime environments, with a set of data processing elements connected in series on remote computing clouds. This program allows analyses to be preserved/reused after its initial commission and not discarded.

REANA sets out to solve issues with version updates, and compatibility issues by containerizing the analyses. REANA's reproducible analysis workshop is provided as a guide to properly setting up and using its platform.

This project will involve a careful inspection of the REANA analysis preservation workshop tutorial, which will involve an examination and testing the efficacy of the workshop itself. This step is already in progress.

Project Proposal/Deliverables:

The project aim of this project is to test REANA and its training material. This study will result in improving the workshop, widening its usability, and mitigating difficult readability.

Timeline:

Week 1 (Full Time)

- Introduce/familiarize myself with REANA.
- Introduce/Familiarize myself with Jekyll and git workflow.

Week 2-Week 7(Full Time)

- ○ Test previous workshops that introduce REANA. Make sure it is a reasonable cross-experiment.
- ○ The material should be verbose enough to be followed for self-study.
- ○ Fix small issues (typos, dead links, etc.) immediately, identify larger issues, and collect them as GitHub issues or in a report.
- ○ Consult with the original authors/training team on how to address larger issues.
- ○ Fix/address issues found. Along with constructing a list of the issues for future reference and maintenance.

References:

[1] *REANA - Reusable Analyses*. https://www.reana.io/. Accessed 21 Apr. 2023.

[2] *Reproducible Analyses with REANA*.

https://hsf-training.github.io/hsf-training-reana-webpage/. Accessed 21 Apr. 2023.

[3] "Training on Analysis Preservation (Virtual)." *Indico*, 16 Jan. 2023,

https://indico.cern.ch/event/1219810/.