

On-Demand Distributed Computing Workflow for Physics Analysis at the CMS Experiment



Diyaselis Delgado¹, Kati Lassila-Perini², Clemens Lange³

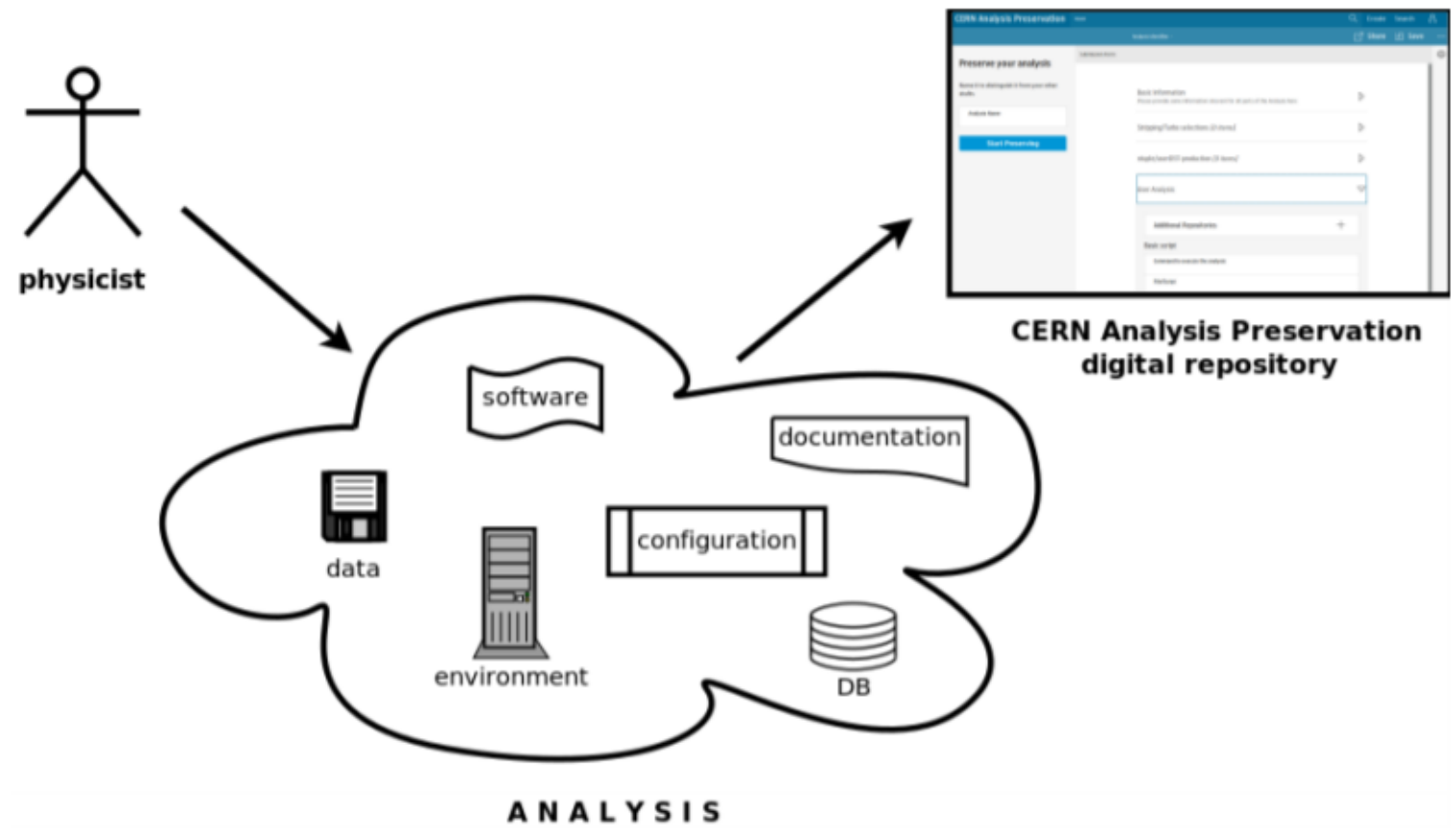
¹University of Puerto Rico Mayaguez, ²Helsinki Institute of Physics, ³European Organization for Nuclear Research

Introduction

The CMS experiment is a strong contributor to the CERN Open Data Portal (CODP) where it publishes research level data together with environment, software and instructions on how to use it. This project takes part in developing and testing simplified examples of open data while providing a connection to analysis preservation and a new reproducible research data analysis platform (ReANA).

CERN Analysis Preservation (CAP)

- The CAP Framework [2] is a central platform for the four LHC collaborations
- Developed to address the need for the long-term preservation of the data analysis process

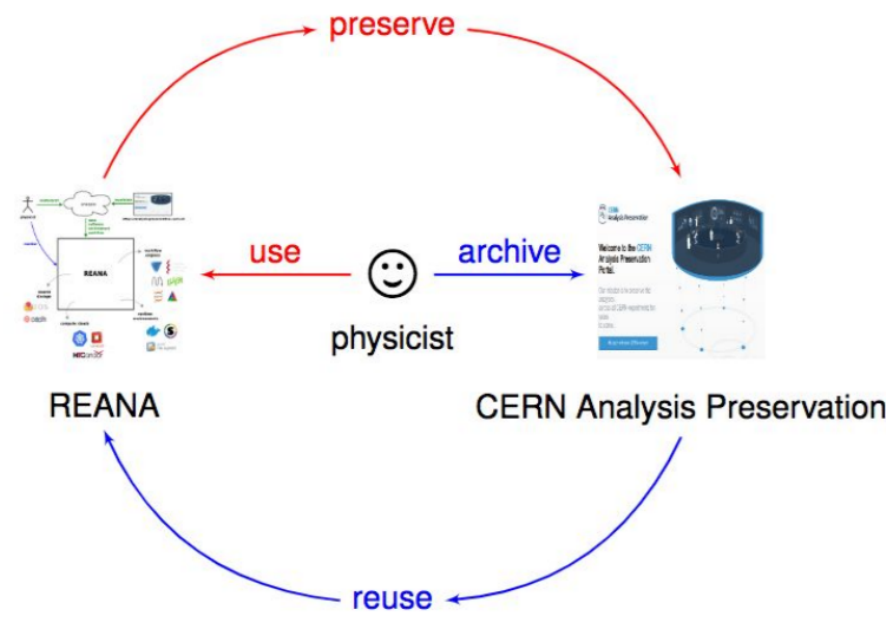


- Capturing knowledge and assets of individual physics analyses in a digital repository in view of facilitating their future reuse.

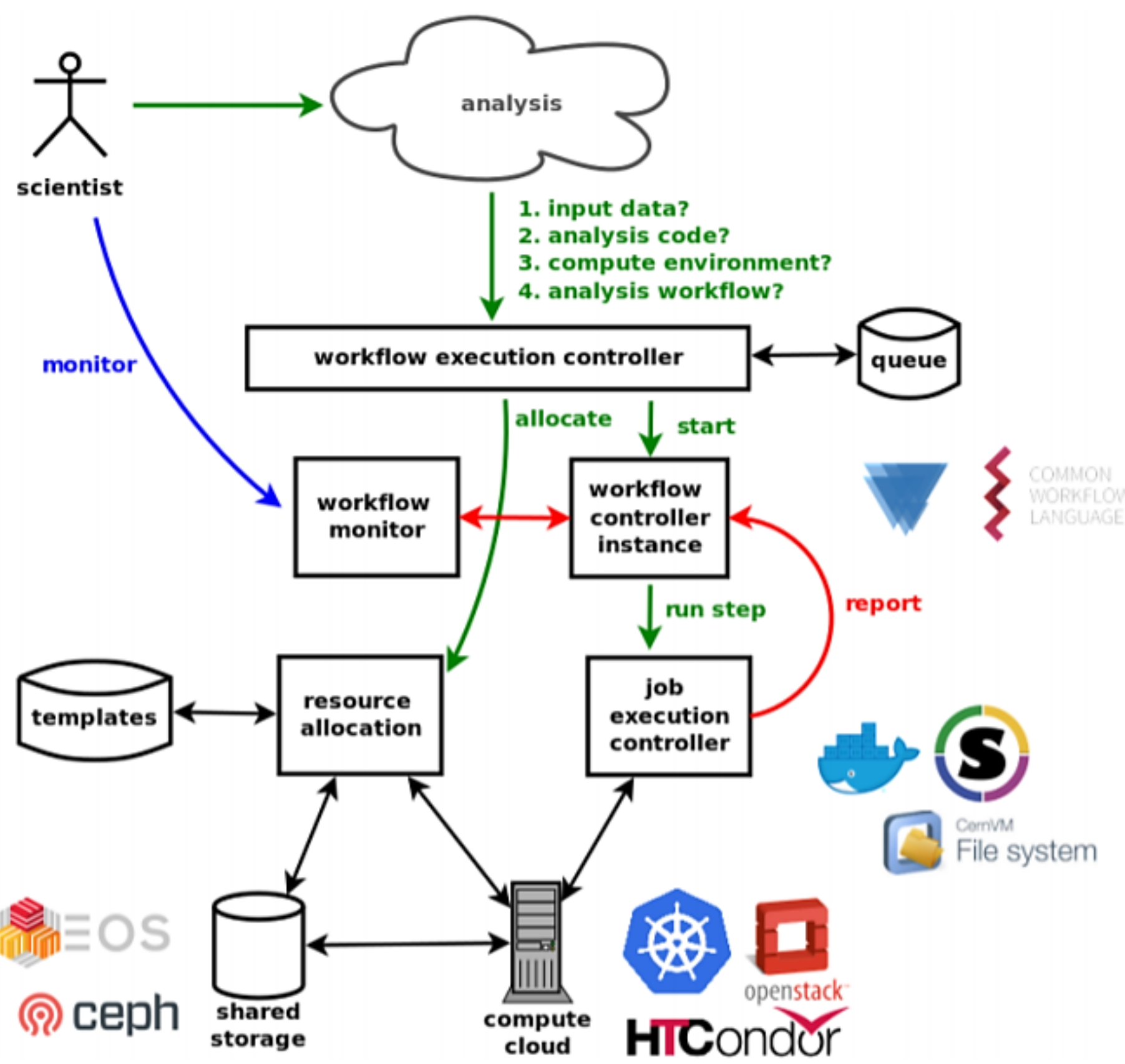
ReANA Platform

reana is a reusable and reproducible research data analysis platform. [3]

- Structures input data, analysis code, containerized environments and computational workflows
- Supports multiple scenarios, such as: computing clouds, running environments, resource orchestration tools, workflow engines, and shared storage systems
- ReANA is applicable to any scientific discipline



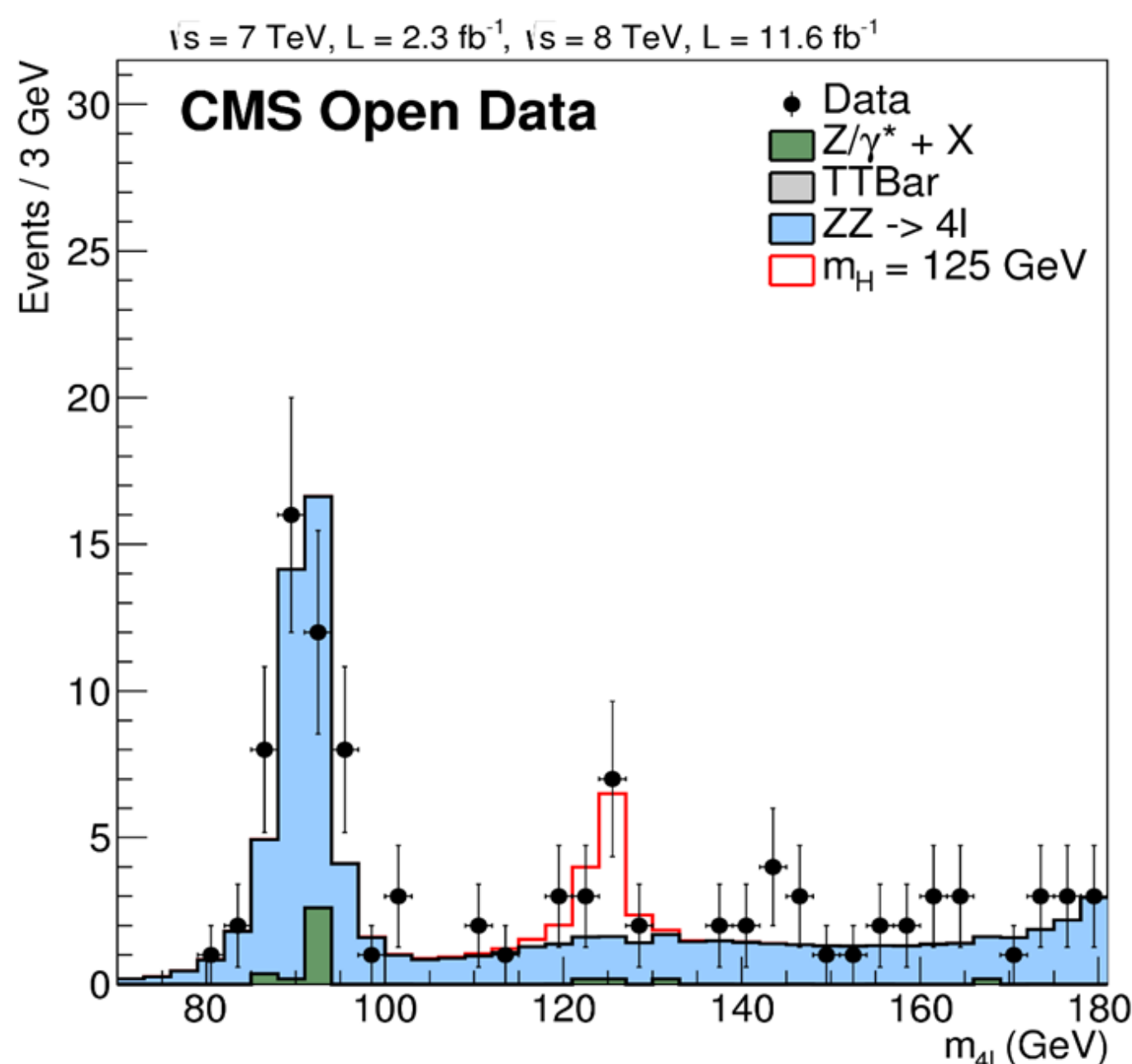
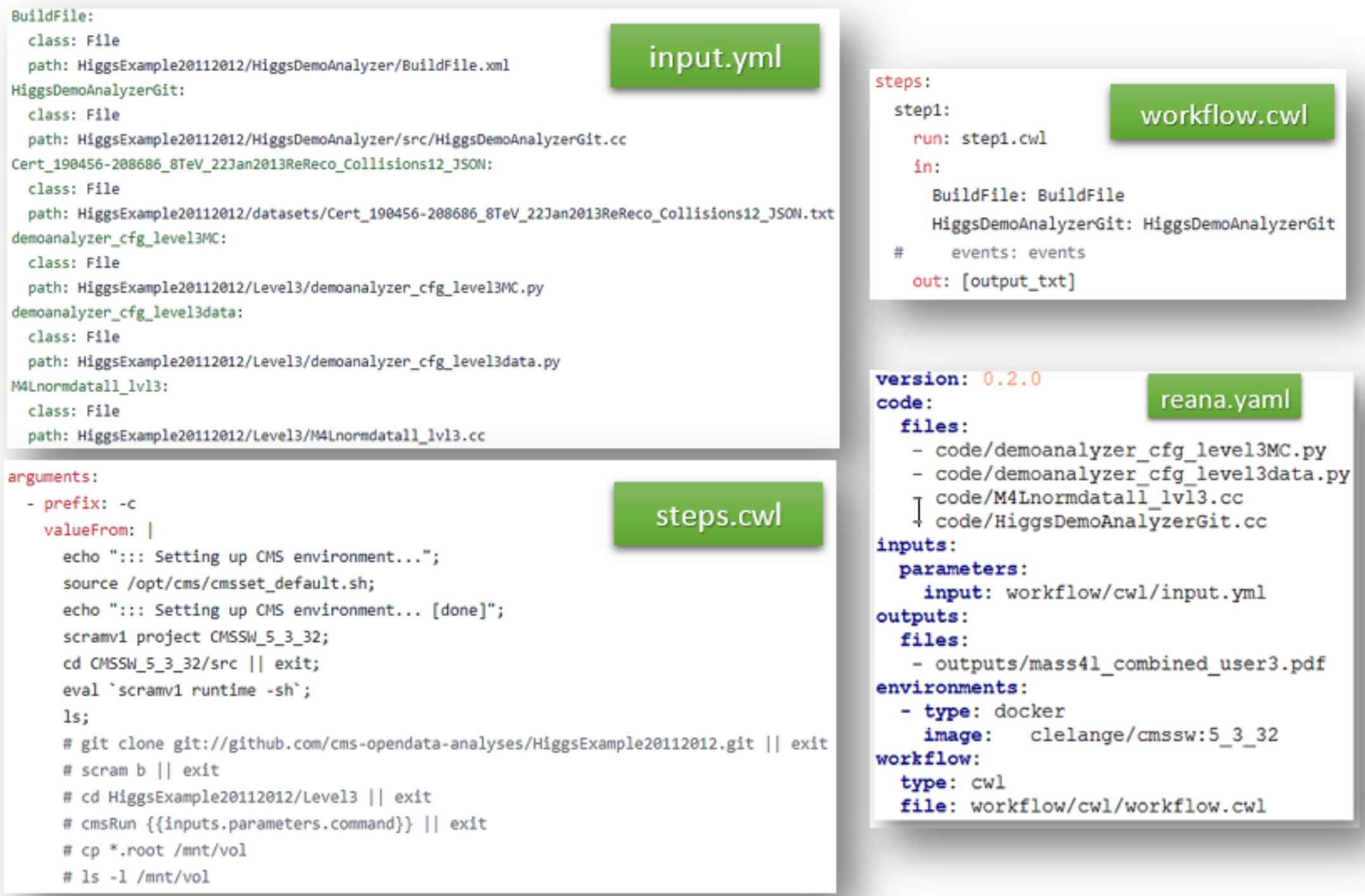
CAP and ReANA Integration



- Building system to instantiate preserved analysis on the cloud

Higgs-to-four-lepton analysis example using 2011-2012 data

- Simplified reimplementaion of the Higgs discovery [1]
- Uses CMS Open Data inputs and available CMS Software environment
- ReANA, captures the analysis workflow and runs the commands to obtain the output



Running the analysis on ReANA

- This CERN project interprets and executes workflows.
- In addition to the existing code, only a set of descriptive workflow files is needed.

```
• Installing the REANA command-line client
$ mkvirtualenv reana-client -p /usr/bin/python2.7
$ pip install reana-client
$ export REANA_SERVER_URL=http://reana.cern.ch
$ eval $(reana-cluster env)
```

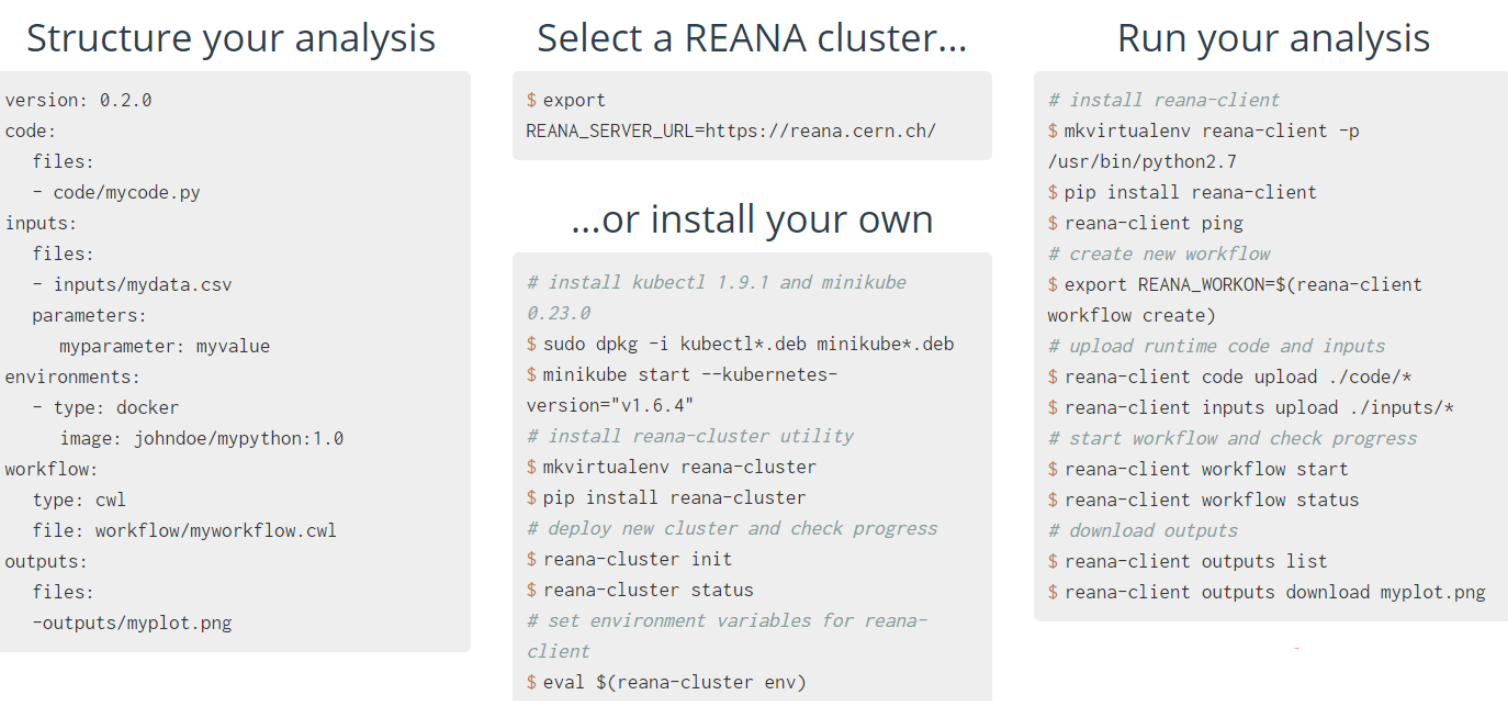
```
• Test client-to-server connection and create a new workflow
$ reana-client ping
Server is running.
$ reana-client workflow create
workflow.1
$ reana-client workflow status -w workflow.1
NAME RUN_NUMBER ID
workflow 1 91797125-012c-498d-8a92-b4f7e3598513
$ export REANA_WORKON=workflow.1
```

- Upload code and start the workflow

CODP Validation Examples

- Validation examples prove that a result can be obtained with the legacy/open data tools
- For example, a validation codes for 2010 datasets [5] consists of:

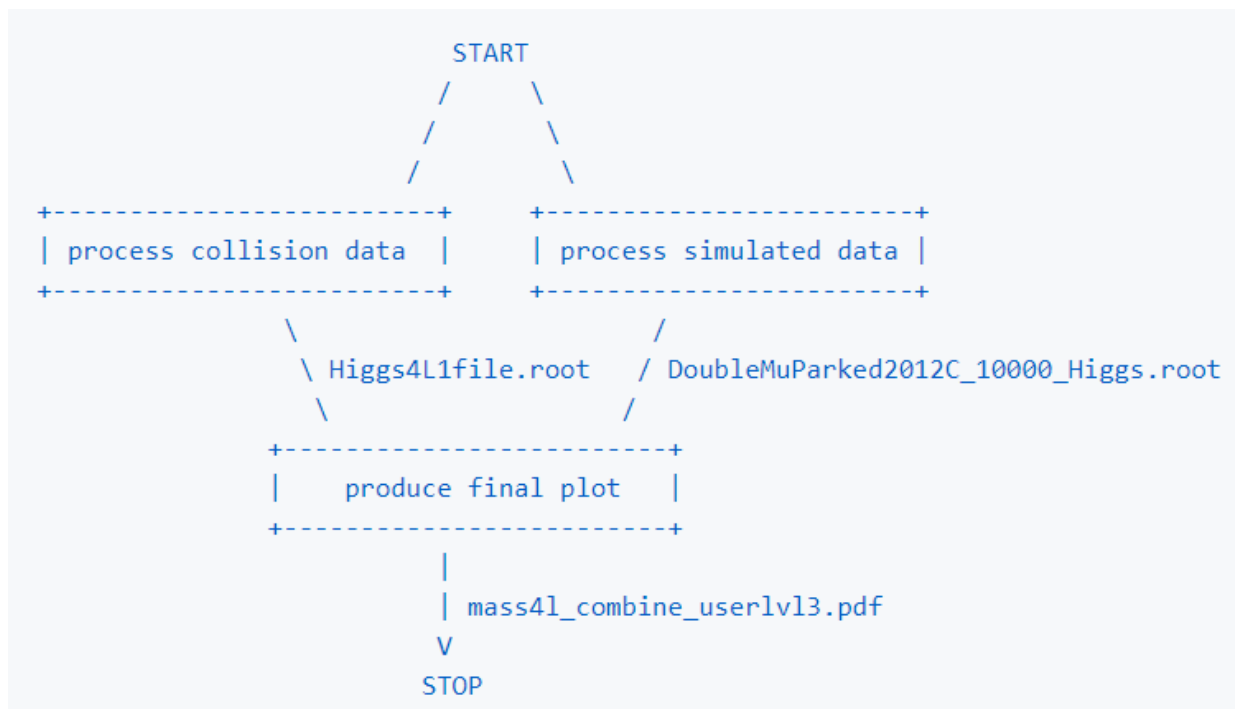
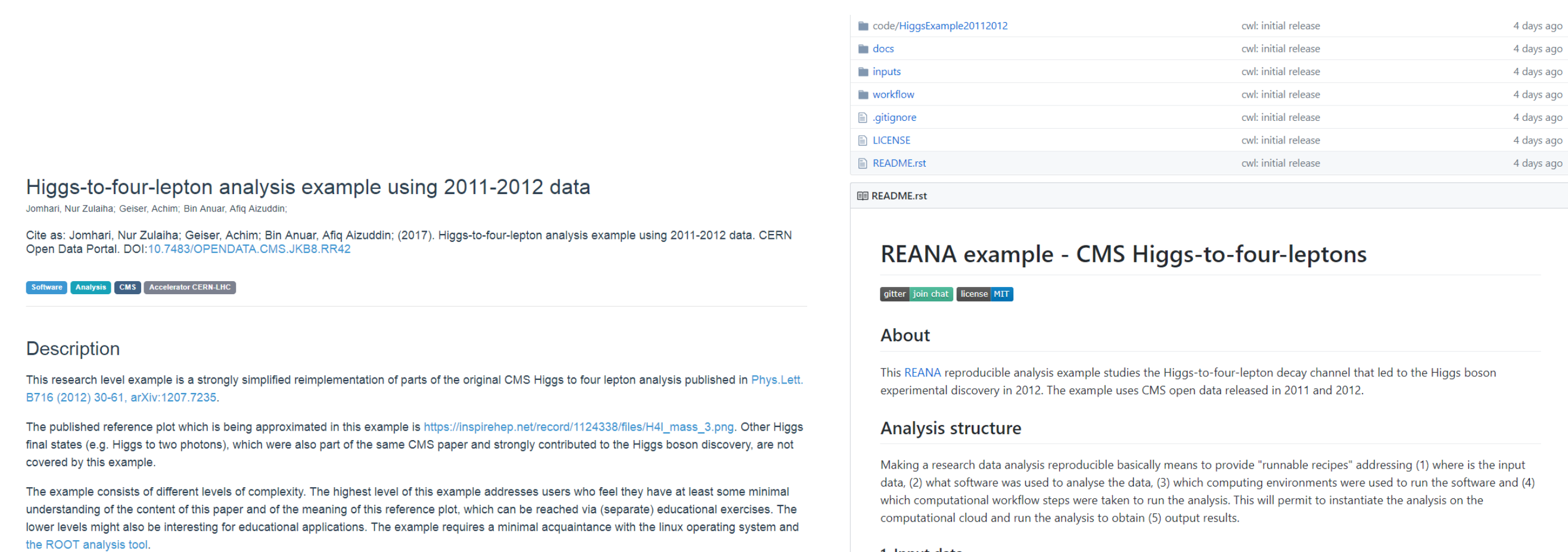
- (1) Setting up the CMS environment
- (2) Compiling and Running the Analysis
- (3) Producing a result (e.g. a comparison plot for the dimuon mass spectrum)



- Demonstrated setup allows to rerun CMS Open Data analyses (e.g. confirm that instructions are correct and run large-scale analyses relatively fast)

Working Example on the Platform

The Higgs-to-four-lepton analysis example has been implemented in ReANA. This CMS Open Data example is up and running. [4]



References

- [1] <http://doi.org/10.7483/OPENDATA.CMS.JKB8.RR42>
- [2] <https://analysispreservation.cern.ch/>
- [3] <http://www.reana.io/>
- [4] <https://github.com/reanahub/reana-demo-cms-h4l>
- [5] <http://opendata.cern.ch/search?page=1&size=20&q=validation%20cms&subtype=Validation>