CERN Open Data
CERN Analysis Preservation
REANA Reproducible Analyses

A brief status

Tibor Simko - CERN IT
Pamfilos Fokianos - CERN SIS

Data & Analysis Preservation Workshop - 26 November 2019

CERN Open Data

Status: *production* (since November 2014)

Size: 7K records, 800K files, 2 PB size

Purpose: “big data” sharing of event-level particle physics data and accompanying code for both education and research purposes

Content: raw samples, collision & simulated & derived datasets, docs, configs, software tools, example analyses, VMs, event display

Community: ALICE, ATLAS, CMS, LHCb, OPERA (coming: JADE, Data Science)

Notes: independent expert curation; batch ingestion workflows with Collaborations

http://opendata.cern.ch

CERN Analysis Preservation

Status: *pilot*

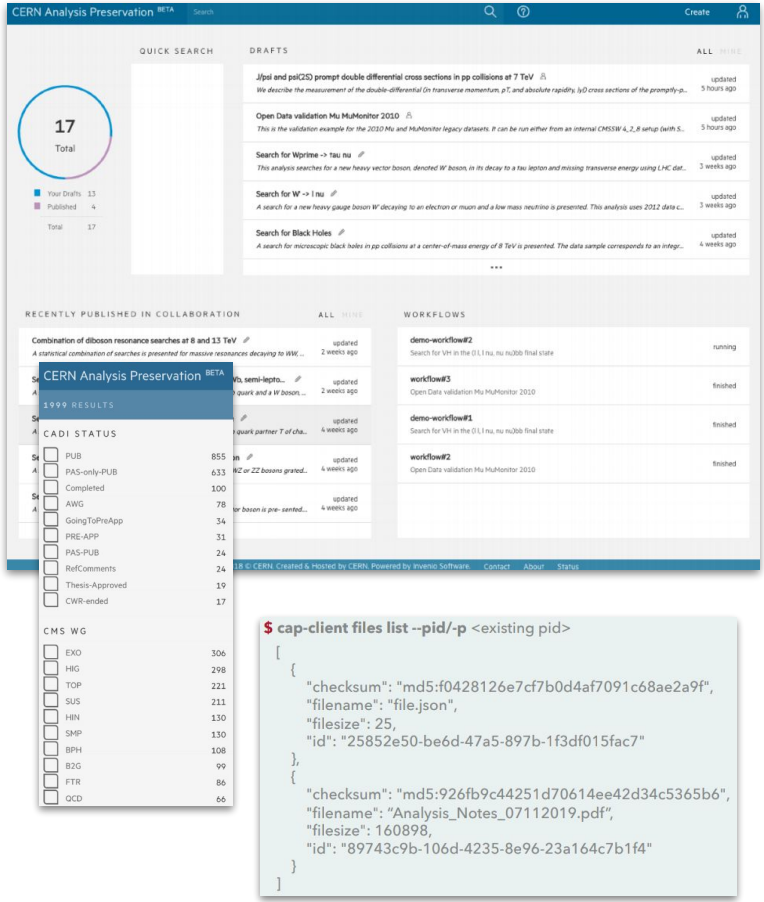
Purpose: safe digital repository for preserving all individual user analysis assets of interest that are not stored anywhere (n-tuples, intermediate data, user code, plotting macros...)

Usage: describe analysis + deposit assets via CLI & Web UI + share with colleagues = preserve knowledge

Community: pilot examples with ALICE, ATLAS, CMS, LHCb; synergies with FREYA EC project partners

Notes: focus on “preservation” of analysis knowledge & assets without necessarily implying “reproducibility”; interconnection with Collaboration DBs and workflows

Resources: shared EOS space



http://analysispreservation.cern.ch

REANA Reproducible Analyses

Status: *pilot*

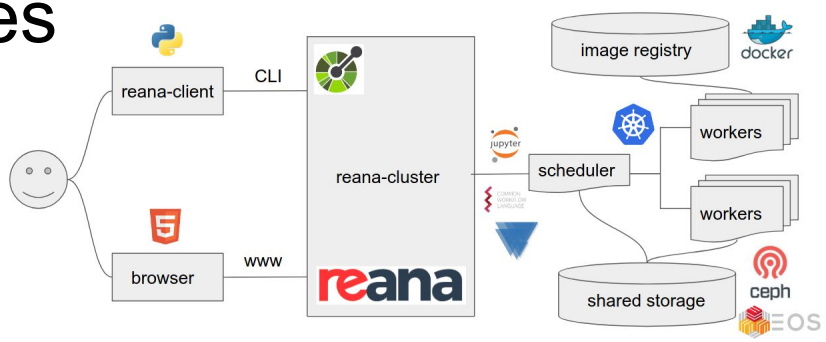
Purpose: run containerised scientific workflows on diverse compute clouds (Kubernetes, Condor, Slurm)

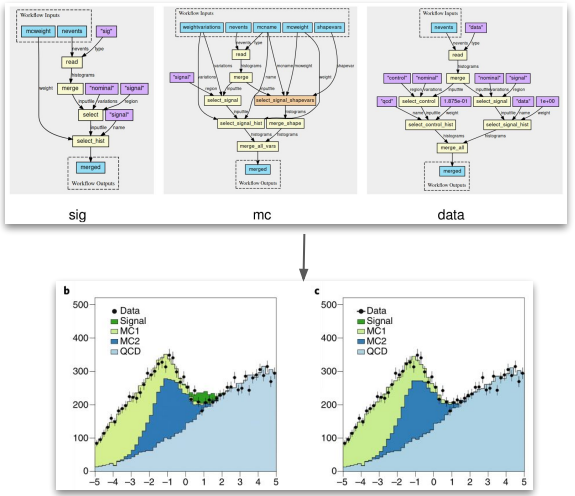
Usage: data + code + environment + workflow = reproducibility

Community: pilot examples with ALICE, ATLAS, CMS, FCC, LHCb; synergies with astronomy, life sciences, machine learning

Notes: focus on “preproducibility” of analysis during its active phase; structure analysis in a reproducible way to facilitate its future “preservation”

Resources: shared Ceph storage and Kubernetes cluster; need for proper experiment accounting





http://www.reana.io

Preservation & Reproducibility