

# CERN Analysis Preservation framework & REANA reproducible analysis platform

Tibor Šimko

CERN

*DPHEP remote preparatory discussion*

*2 March 2021*

<https://indico.cern.ch/event/1009487/>

# CERN Analysis Preservation framework

**Purpose:** capture and preserve all elements needed to understand and reuse an analysis even several years later; take a consistent snapshot linking all the knowledge

**Usage:** describe analysis + deposit n-tuples, code etc via CLI and web UI + share with colleagues = preserve knowledge

**Community:** pilot with ALICE, ATLAS, CMS, LHCb

- ▶ content restricted to collaborations
- ▶ metadata interconnected with collaboration databases
- ▶ associated knowledge, e.g. CMS statistics questionnaire
- ▶ helps addressing increasing number of funding agencies asking for comprehensive data management policies
- ▶ run by CERN Scientific Information Service (P. Fokianos, K. Naim)

The screenshot shows the CERN Analysis Preservation web interface. At the top, there's a search bar and a 'DRAFTS' section with a '17 Total' indicator. Below that, there are search filters and a 'RECENTLY PUBLISHED IN COLLABORATION' section. A sidebar on the left contains filters for 'CADI STATUS' and 'CMS WG'. A terminal window in the bottom right shows the command 'cap-client file list --pid/p <existing pid>' and its output, which lists file metadata including checksum, filename, file size, and ID.

<https://analysispreservation.cern.ch>

# REANA reproducible analysis platform

**Purpose:** run declarative computational workflows on containerised compute clouds

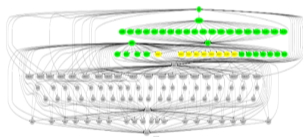
**Usage:** data + code + environment + workflow = computational reproducibility

**Community:** pilot examples with ALICE, ATLAS, CMS, FCC, LHCb; ATLAS search groups (SUSY, EXOT, HDBS) now require workflow preservation as mandatory for analysis approval

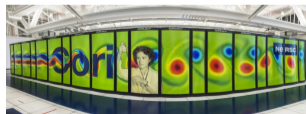
- ▶ promotes pre-reproducibility during active analysis phase to facilitate future preservation
- ▶ integration with GitLab; CI/CD mode
- ▶ verification of analysis examples and data provenance chain (CMS AOD reprocessing)
- ▶ support for hybrid compute workflows with multiple backends (HTCondor, Kubernetes, Slurm)



<https://www.reana.io>



CMS Jet Energy Corrections workflow



REANA running on supercomputers (e.g. NERSC)