

Use Cases for the EOSC Federation

24 June 2026, 15:00-16:30



Presenter:

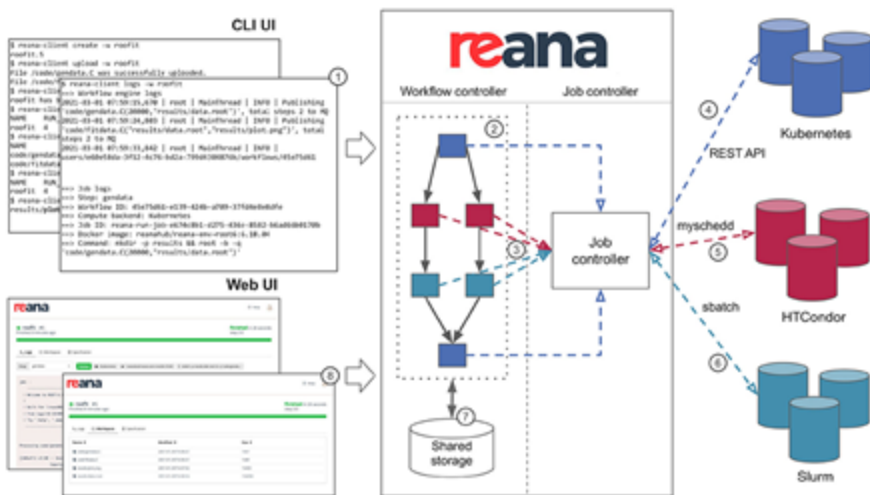
Tibor Šimko

CERN

EOSC Node Coordinator Deputy



- REANA is a reusable and reproducible data analysis platform. Core philosophy:
 - offer a containerised platform allowing researchers to organise and execute their declarative scientific workflows on containerised compute clouds
 - offer multiple workflow engines to fit popular user habits in the community



Multiple compute backends:

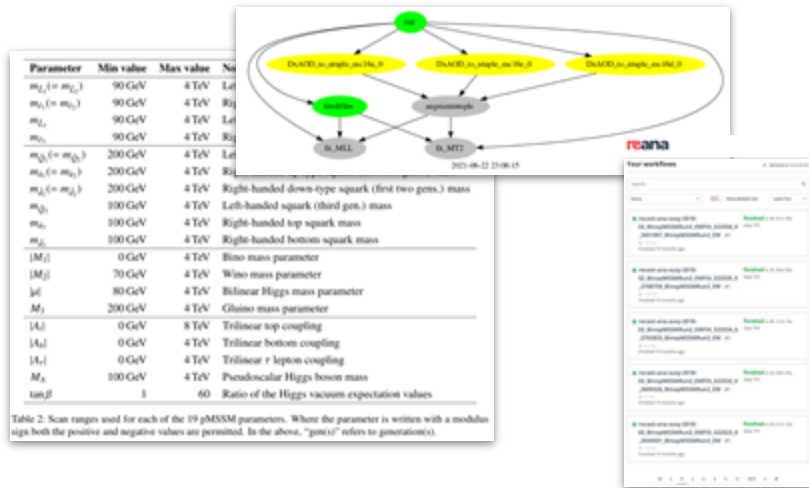
- Kubernetes, HTCondor, Slurm

Multiple workflow languages:

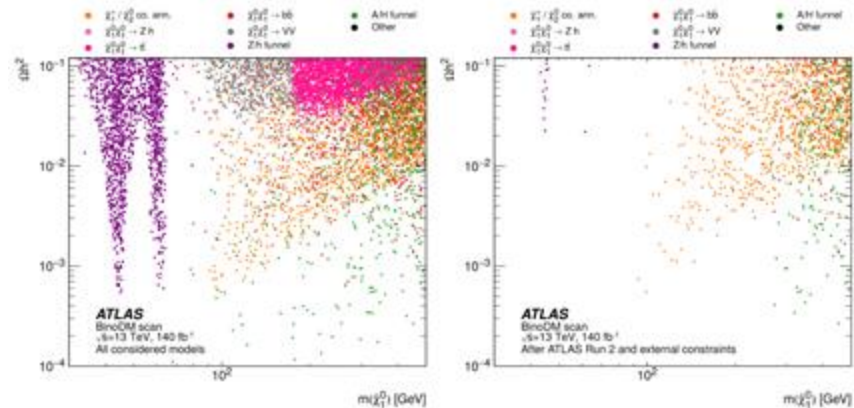
- CWL, Dask, Serial, Snakemake, Yadage

Multiple means of use:

- Command-line client, web interface

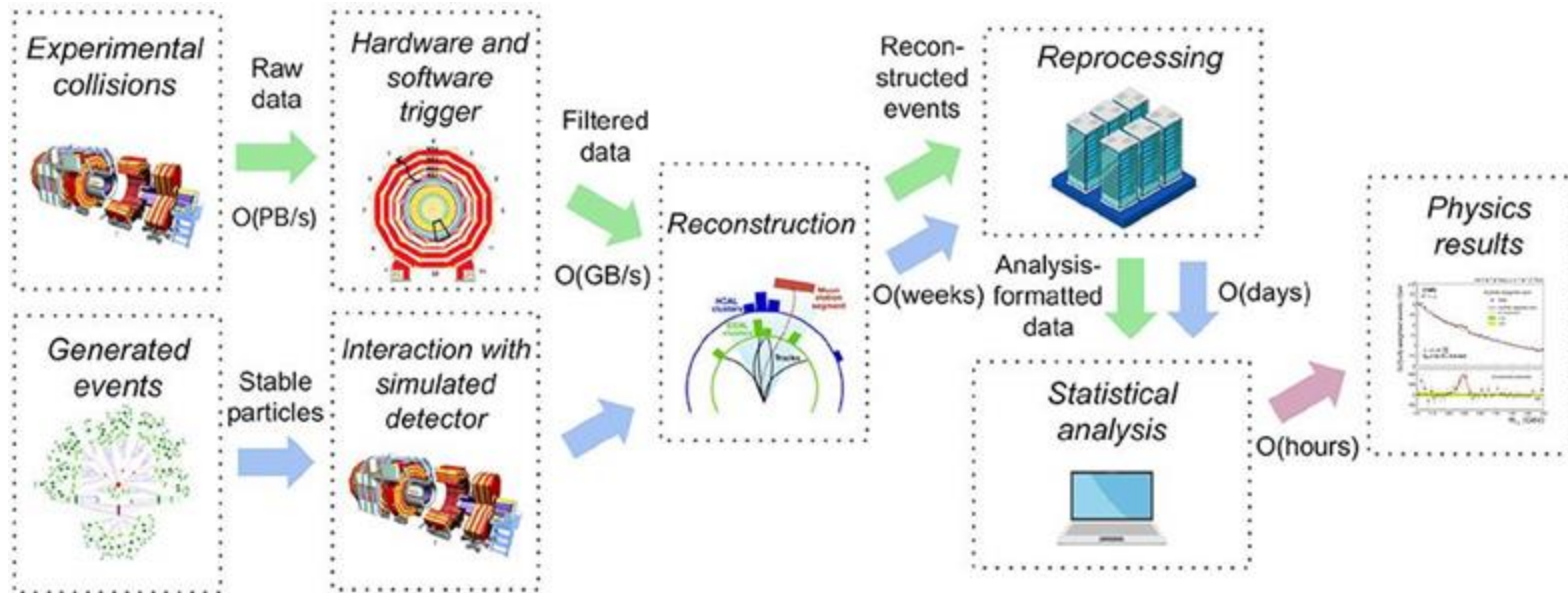


The pMSSM model parameters and a typical computational workflow. Thousands of workflow runs are necessary to cover the parameter space. ([src](#))



The distribution of unexcluded supersymmetry models as a function of two model parameters involving dark matter, shown before (left) and after (right) the results of the 2023 ATLAS pMSSM analysis. ([src](#))

Original use case: facilitate searches for new physics beyond standard model



Typical experimental particle physics data production and analysis workflow ([src](#))



CMS 2012 collision data



CMS 2012 simulated data



CMS 2012 validated runs



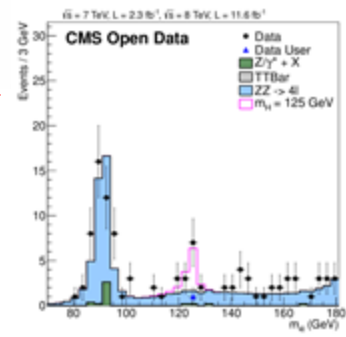
workflow



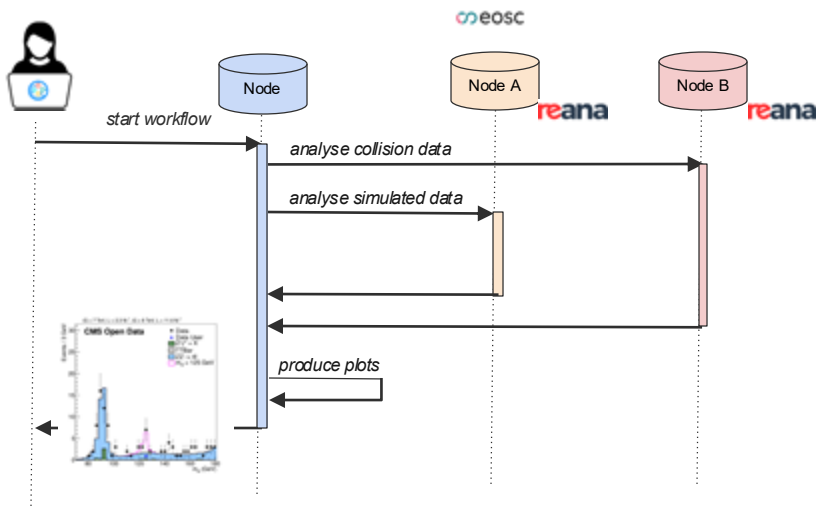
container



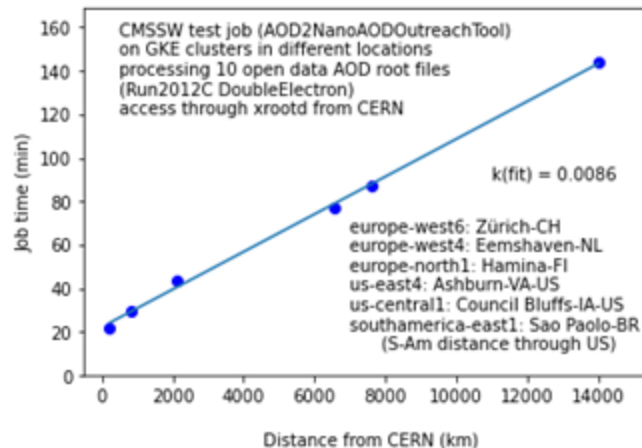
code



CMS open data H→4l analysis example ([src](#))



The use case goal: allow dispatching parts of computations (1) to Nodes where the data sits; (2) to Nodes where the user has computing resources.



An example of importance of near-data computing: the performance of CMS AOD data processing on various near and far public clouds ([src](#))

- The use case addresses two kinds of federating computing needs
 - *Near-data computing*, to send computational workflow jobs to where large scientific datasets are stored, rather than transferring massive volumes of data to the researcher. Example: processing datasets stored on different Nodes.
 - *Closed-data computing*, to allow external collaborators to reinterpret existing analyses without offering direct data access. Example: test new ideas of theorists who do not have access to the original experimental data (ATLAS RECAST).
- The target user group:
 - Researchers writing declarative computational analysis workflows (any scientific discipline)
 - Infrastructures wanting to offer remote compute capabilities to researchers

- During the EOSC Federation build-up phase, the following EOSC Nodes participated in the use case demonstrator:
 - EOSC CERN Node
 - EOSC EU Node
 - EOSC Finland
 - EOSC Italy
 - EOSC Poland
 - EOSC SURF Netherlands
- The cross-Node collaboration required technical coordination of federated compute solutions (Kubernetes with MultiKueue)

- The core EOSC Federation capabilities used
 - FC-1 AAI
 - Authenticating users and acting “on behalf” of users
 - FC-4 Service Monitoring
 - Ensure workloads over collaborating clusters
 - FC-6 Service Accounting
 - Ensure user quota accounting across sites
 - FC-9 Application Deployment Management
 - Payload execution across several Nodes

- A success criteria from a researcher point of view
 - As a researcher, I want to run a declarative analysis on one Node, and see it dispatch jobs to various other collaborating Nodes based on my provided workflow hints
 - As an infrastructure provider, I want to configure and monitor user payloads coming from remote collaborating Nodes
- A demonstrator using CMS open data Higgs to four lepton analysis example was implemented and successfully run across the participating Nodes.
- The developed solutions for generic workflow engines (CWL, Snakemake) are directly applicable in other scientific disciplines (astronomy, life sciences)

- The challenges were mostly technical
 - how to dispatch computing payload across different Kubernetes clusters?
 - how to prevent “unsafe” computing payloads?
- Future improvements
 - efficient propagation of produced data and workflow workspaces (Rucio)
 - see also: FORGE “Federating Open Research data for Genomics on EOSC” with EOSC CERN Node, EOSC Italy, AREA Science Park

Stay in contact!



<https://eosc.eu/academy>



academy@eosc.eu

