eosc | Federation

CERN

CSC

EOSC Node | European Commission

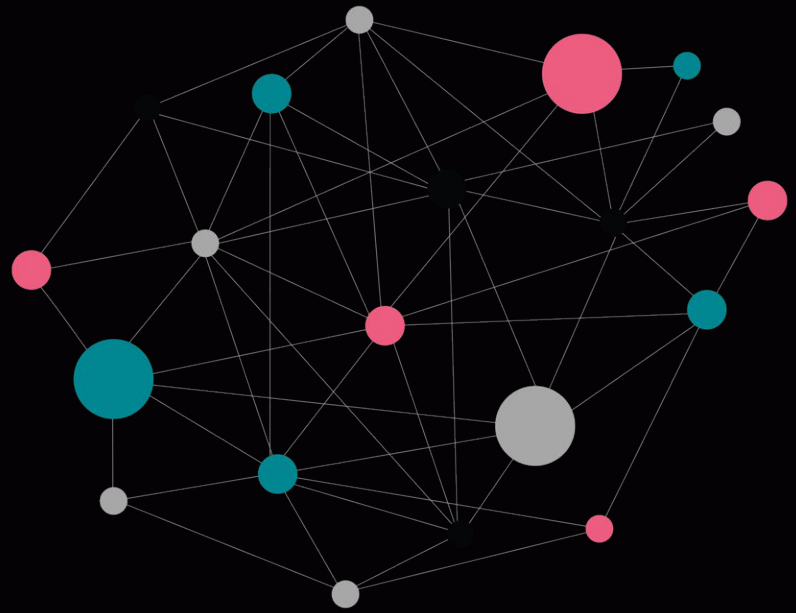
Fondazione ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

National Science Centre

SURF

The REANA use case in the EOSC Federation

FAIR (re)analysis of the LHC data in a distributed environment

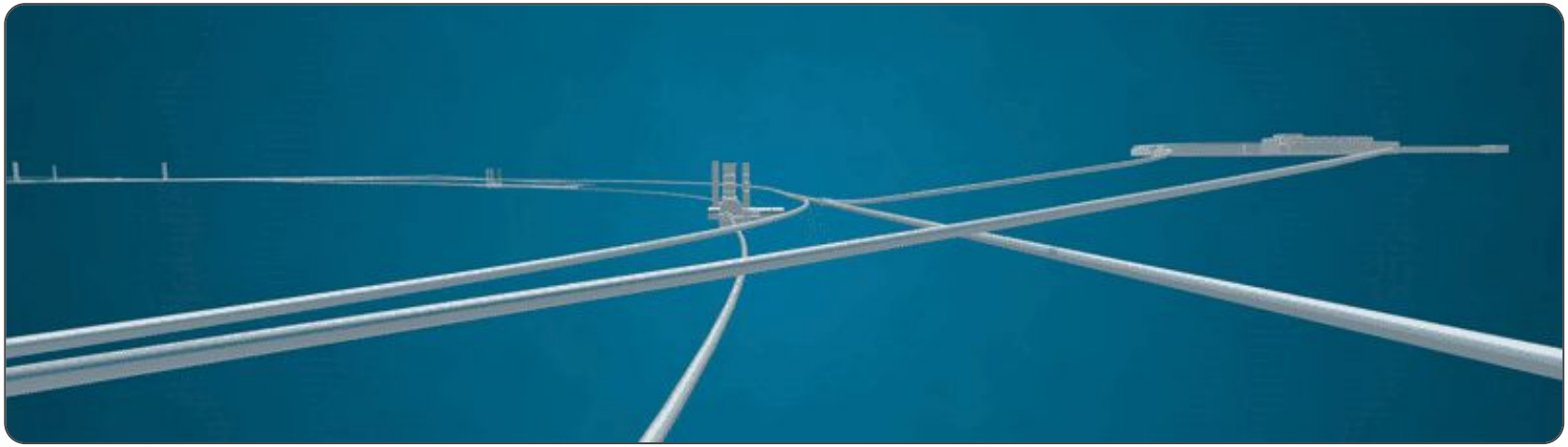


Cameron McClymont (CERN), on behalf of Tibor Simko

22 October 2026

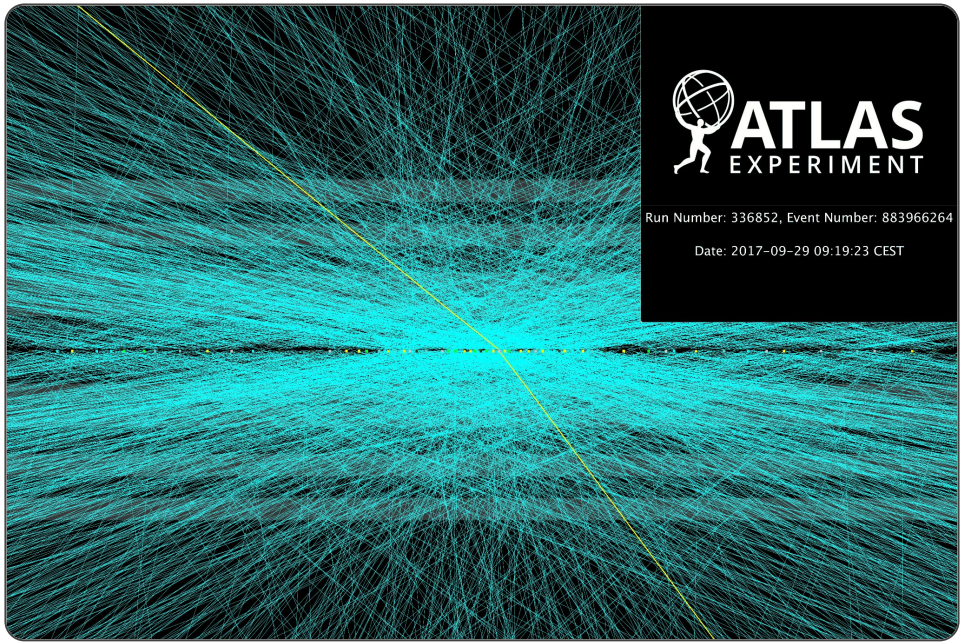
The Large Hadron Collider (LHC)

- The LHC at CERN accelerates (not only) **protons** close to the speed of light
- Bunches of protons **collide** at the interaction points of the detectors



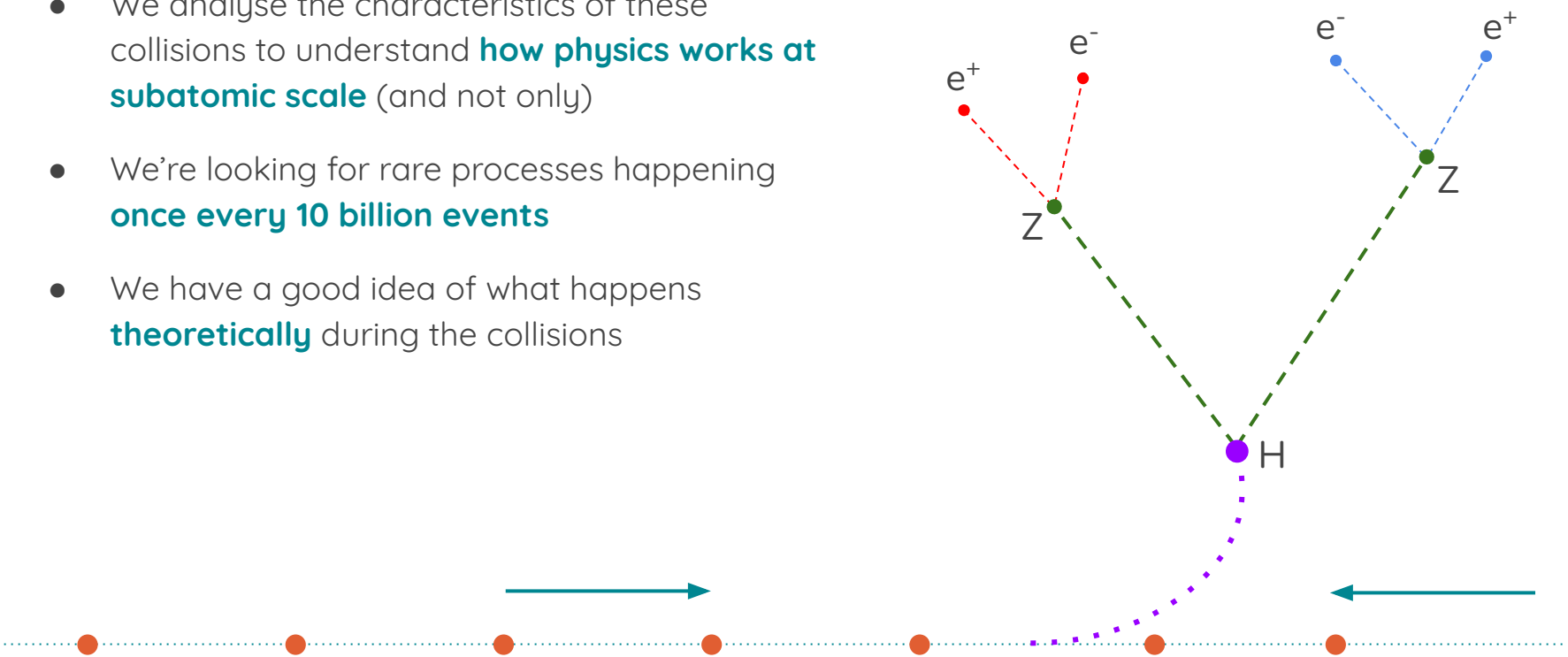
The Large Hadron Collider (LHC)

- The ATLAS and CMS detectors record **40 million proton-proton collisions (events) per second**
- Each collision is a complex mix of thousands of **new particles** being produced



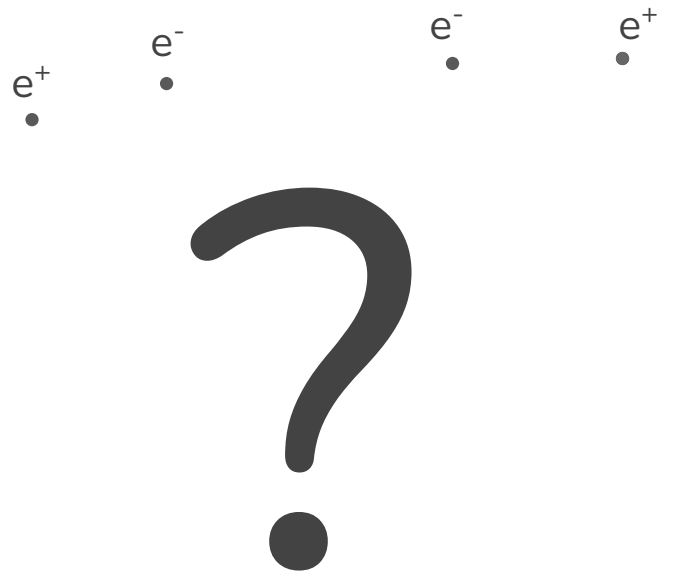
The Large Hadron Collider (LHC)

- We analyse the characteristics of these collisions to understand **how physics works at subatomic scale** (and not only)
- We're looking for rare processes happening **once every 10 billion events**
- We have a good idea of what happens **theoretically** during the collisions



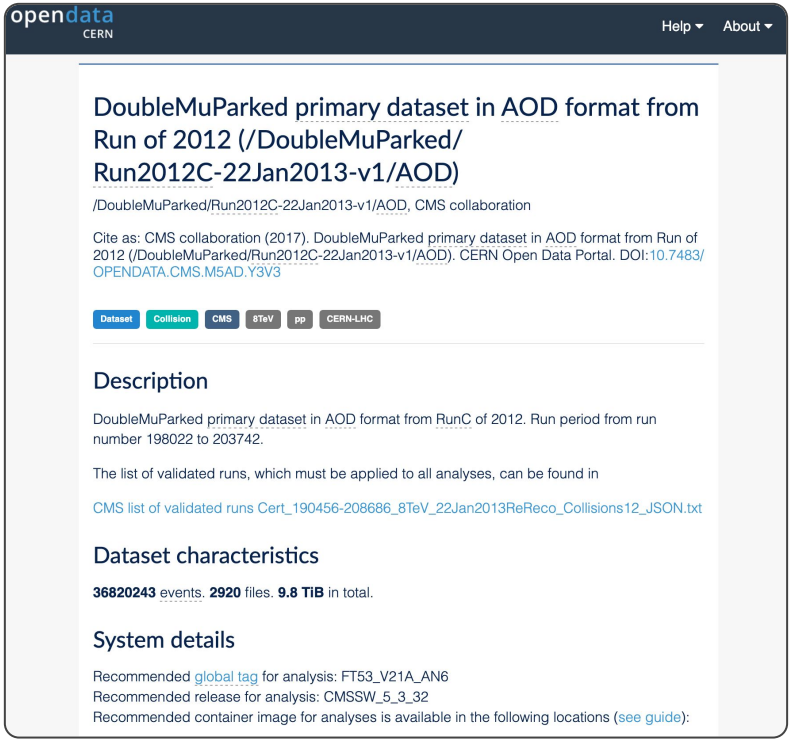
The value of the data

- In practice, the detector only records a **part** of the physics interaction
- We then need to **simulate** the events (according to our theory), and compare them with the output of our detectors



The value of the data

- Each experiment simulates up to ~200B events per year (O(10)PB of data per year), to be used for publications
- CERN currently exposes more than **5 PB** of Open Data
  - Collision data: **unique, only producible at the LHC**
  - Simulated data: produced at a scale (WLCG) that is **difficult to reproduce** by most of the external communities



Questions from researchers

1. How do I run my analysis on a dataset that is **too big** to be moved, or that I **don't have direct access** to?

and / or

2. How do I run my analysis in a place that has **enough computing capacity** to process the data?

Containerised data analyses

https://www.reana.io

reana

Reproducible research data analysis platform

Flexible

Run many computational workflow engines.

COMMON WORKFLOW LANGUAGE

Scalable

Support for remote compute clouds.

kubernetes HTCondor slurm workload manager

Reusable

Containerise once, reuse elsewhere. Cloud-native.

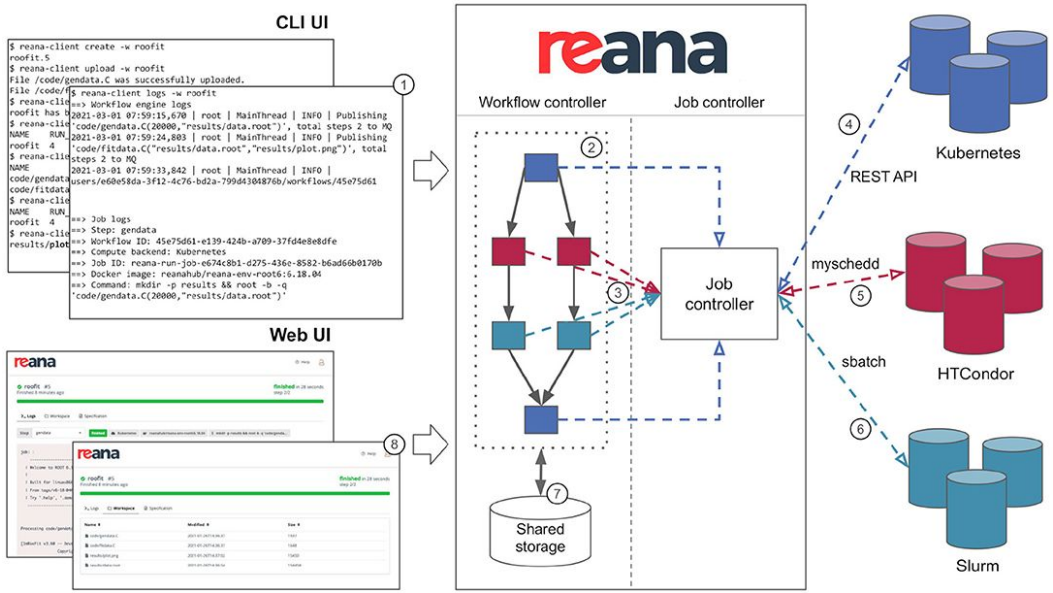
Free

Free Software. MIT licence. Made with ❤ at CERN.

CERN

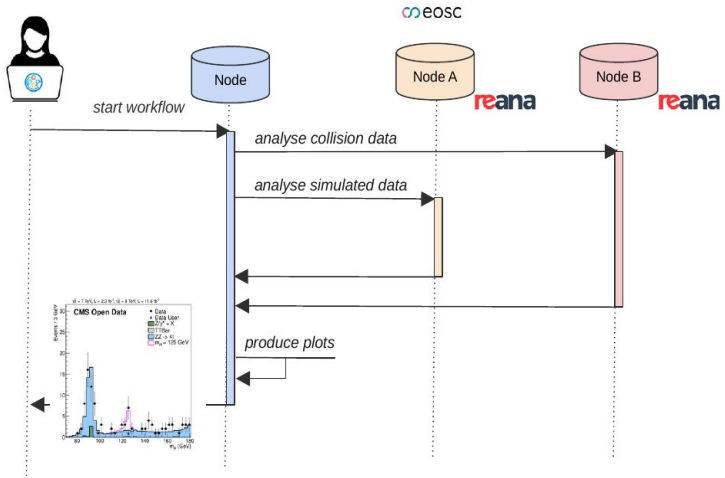
REANA architecture

- Platform for **declarative workflows**, executed in containers
- **Interfaces**: command line, web
- **Backends**: Kubernetes, HTCondor, Slurm
- **Languages**: CWL, Dask, Serial, Snakemake, Yadage



Near-data computing

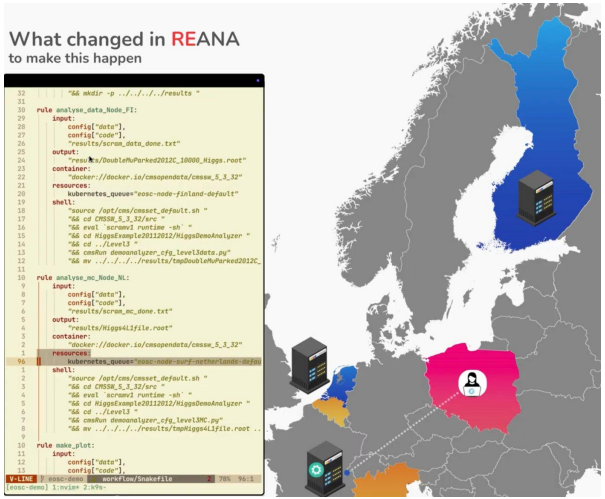
- **Send the job to where the data sits**, not the data to the researcher (infeasible to transfer the large amounts of data needed for some analyses)
- Only send the results back to the researcher to be used in the rest of the workflow



Figures: T. Šimko, EOSC Academy, 24 June 2026

Cross-Node collaboration

- Six EOSC Nodes in the demo: CERN, EU, Finland, Italy, Poland, SURF Netherlands
- Kubernetes with **MultiKueue** dispatches jobs across clusters
- A custom **Snakemake workflow hint** steers each job to a Node
- REANA gathers the results to continue the workflow





Figures: T. Šimko, EOSC Academy, 24 June 2026

EOSC Federation capabilities

- **FC-1 AAI**
  - Authenticate users and act on their behalf
- **FC-4 Service Monitoring**
  - Ensure workload availability over collaborating sites
- **FC-6 Service Accounting**
  - Ensure user quota accounting across collaborating sites
- **FC-9 Application Deployment Management**
  - Execute containerised payload across collaborating sites

Challenges and lessons learned

- **Success**: the H→4l demo ran across the participating Nodes
  - Researcher: run on one Node, jobs dispatched by workflow hints
  - Provider: configure and monitor payloads from remote Nodes
- **Challenges**: mostly technical
  - Dispatching payloads across different Kubernetes clusters
  - Preventing potentially unsafe payloads
- **Next**: standardise data movement between nodes
  - Efficient propagation of produced data and workflow workspaces (Rucio)
  - FORGE: genomics on EOSC, with CERN Node, EOSC Italy and AREA Science Park

REANA meets the EOSC Federation

- Enabled REANA to run federated execution of workflows *à la WLCG* on EOSC Federation
  - **Near-data computing** to efficiently process large amounts of unique data
- Support a coherent, adaptable approach to research
  - Limits the **duplication** of infrastructures
  - Encourages diverse communities to **collaborate**, share expertise, and address common challenges **together**.
- Not only for particle physics!
  - Anyone that needs to run distributed workflows can leverage these functionalities
  - Run any CWL, Snakemake, Yadage workflows in **any scientific discipline**



Thank you!





EOSC Node | European Commission





