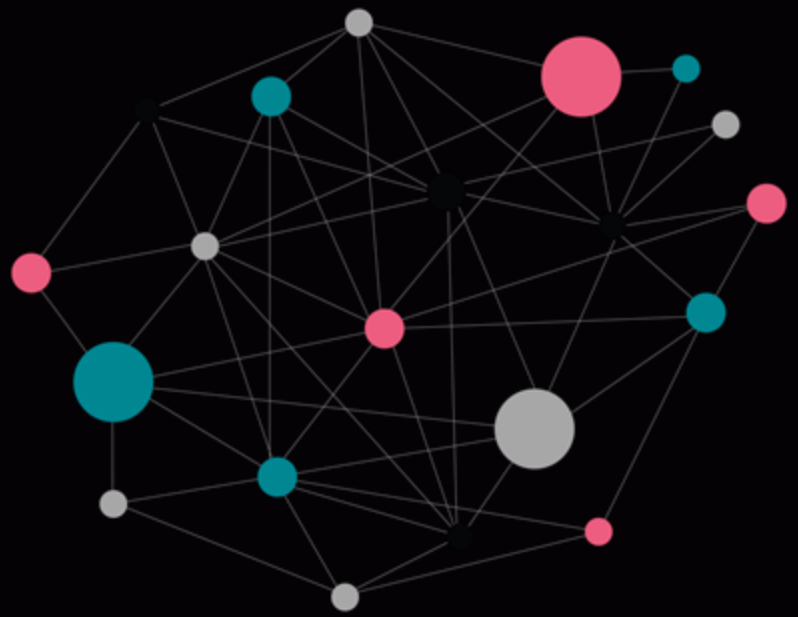
The REANA use case in the EOSC Federation

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

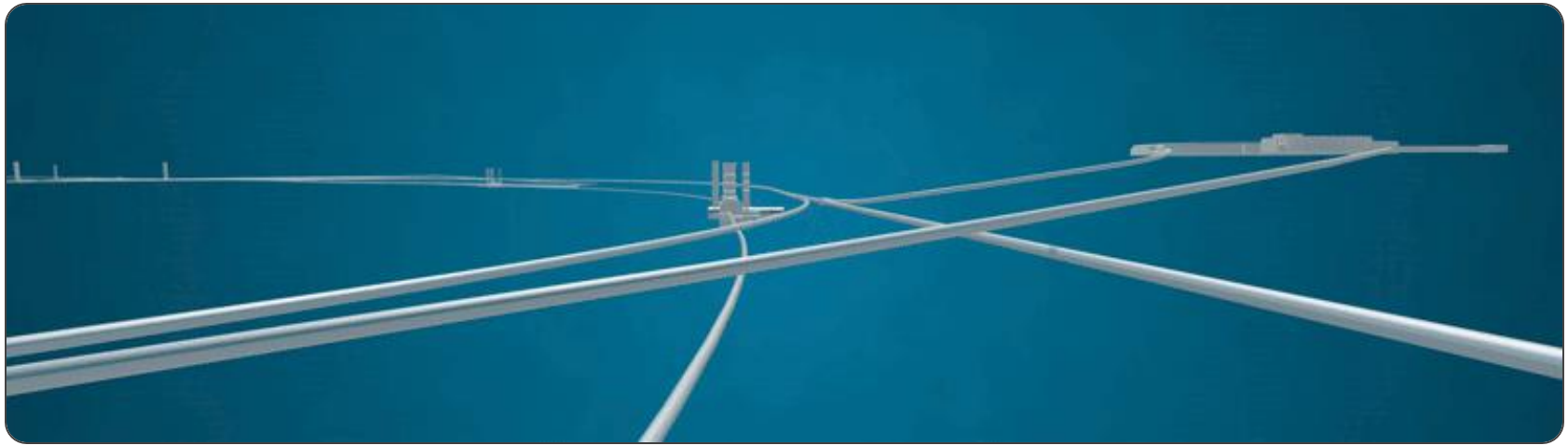


Giovanni Guerrieri (CERN), on behalf of Tibor Simko and the EOSC Scientific Workflows subgroup

3 November 2025

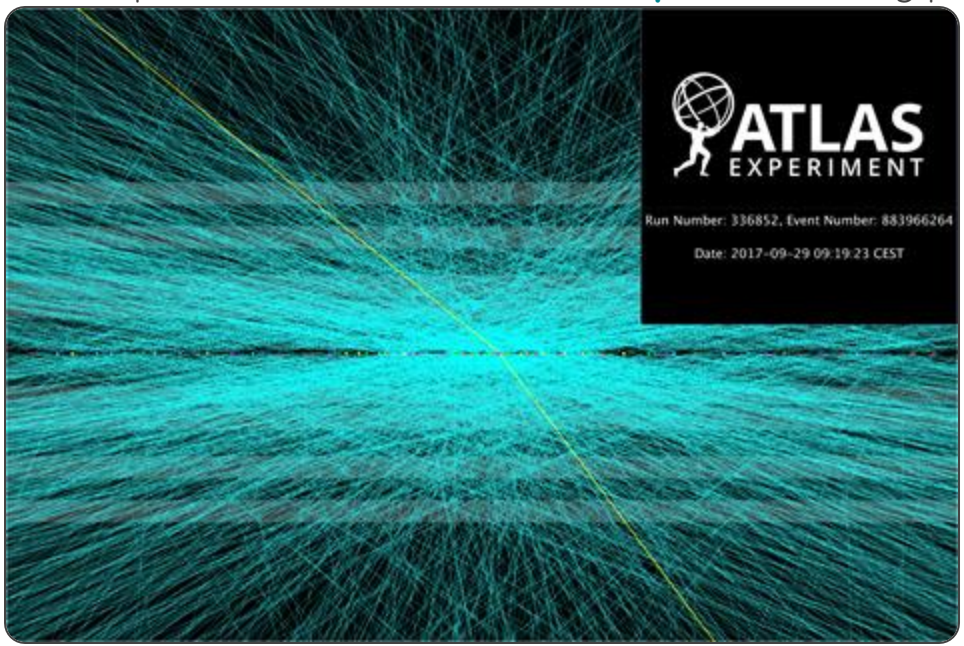
The Large Hadron Collider (LHC)

- 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



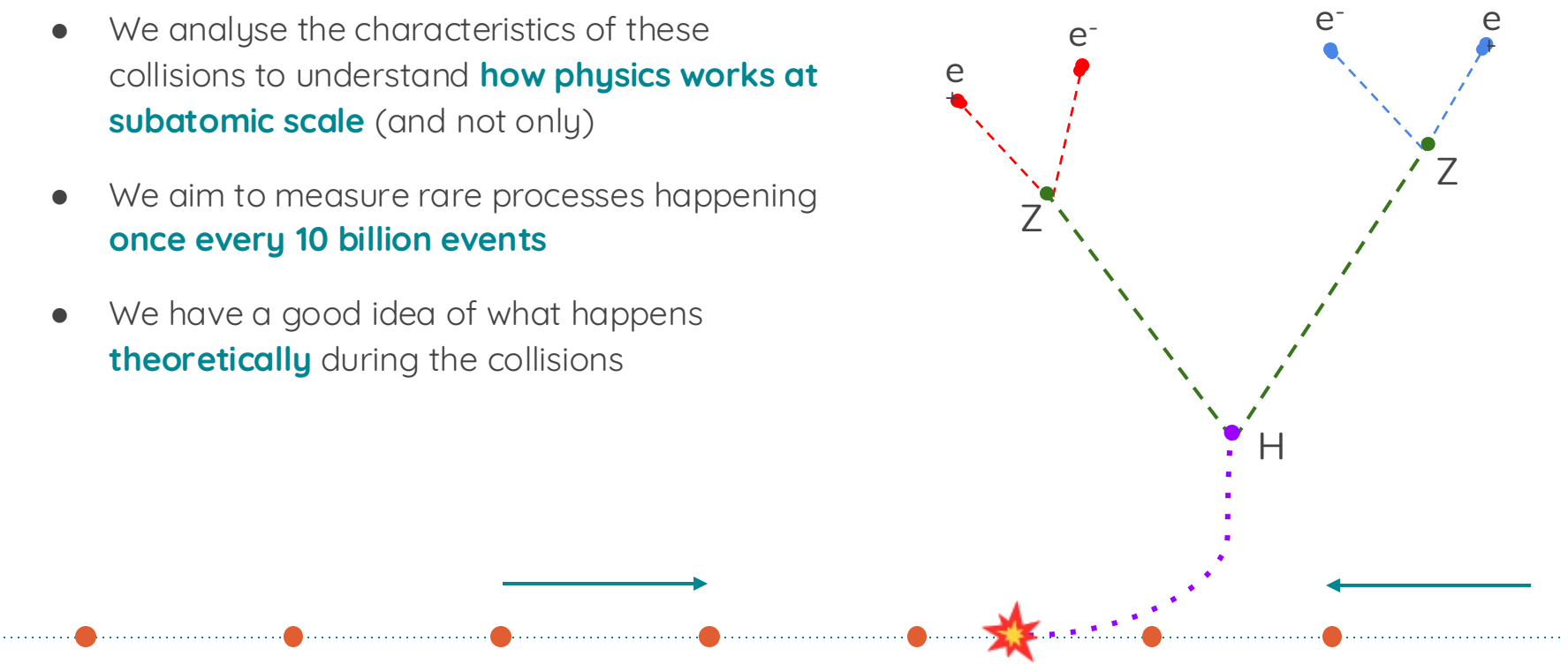
The Large Hadron Collider (LHC)

- 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



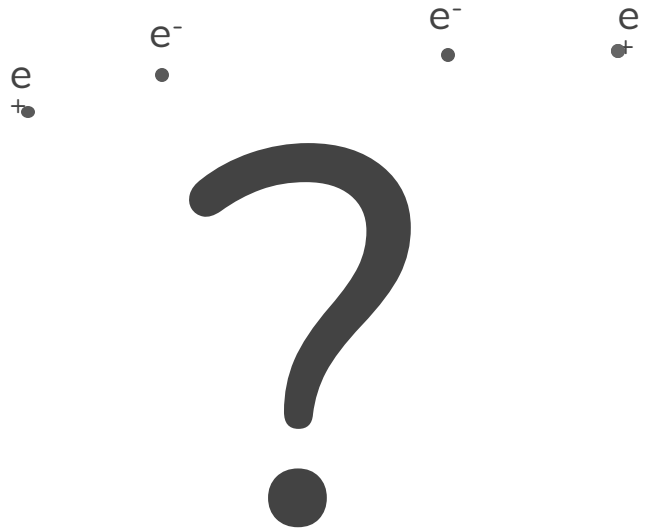
The Large Hadron Collider (LHC)

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



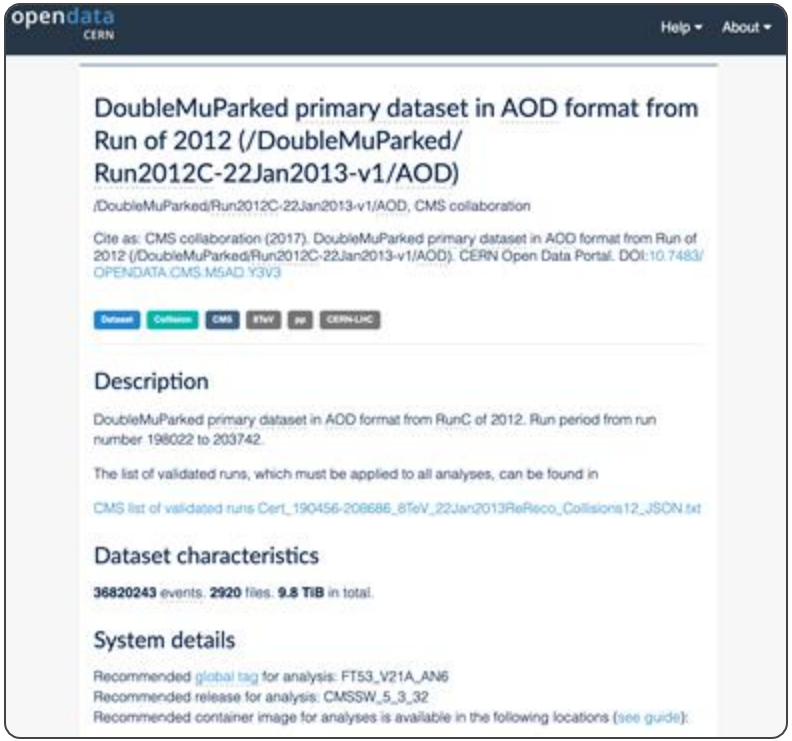
The value of the data

- 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 (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



Frequently asked questions

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

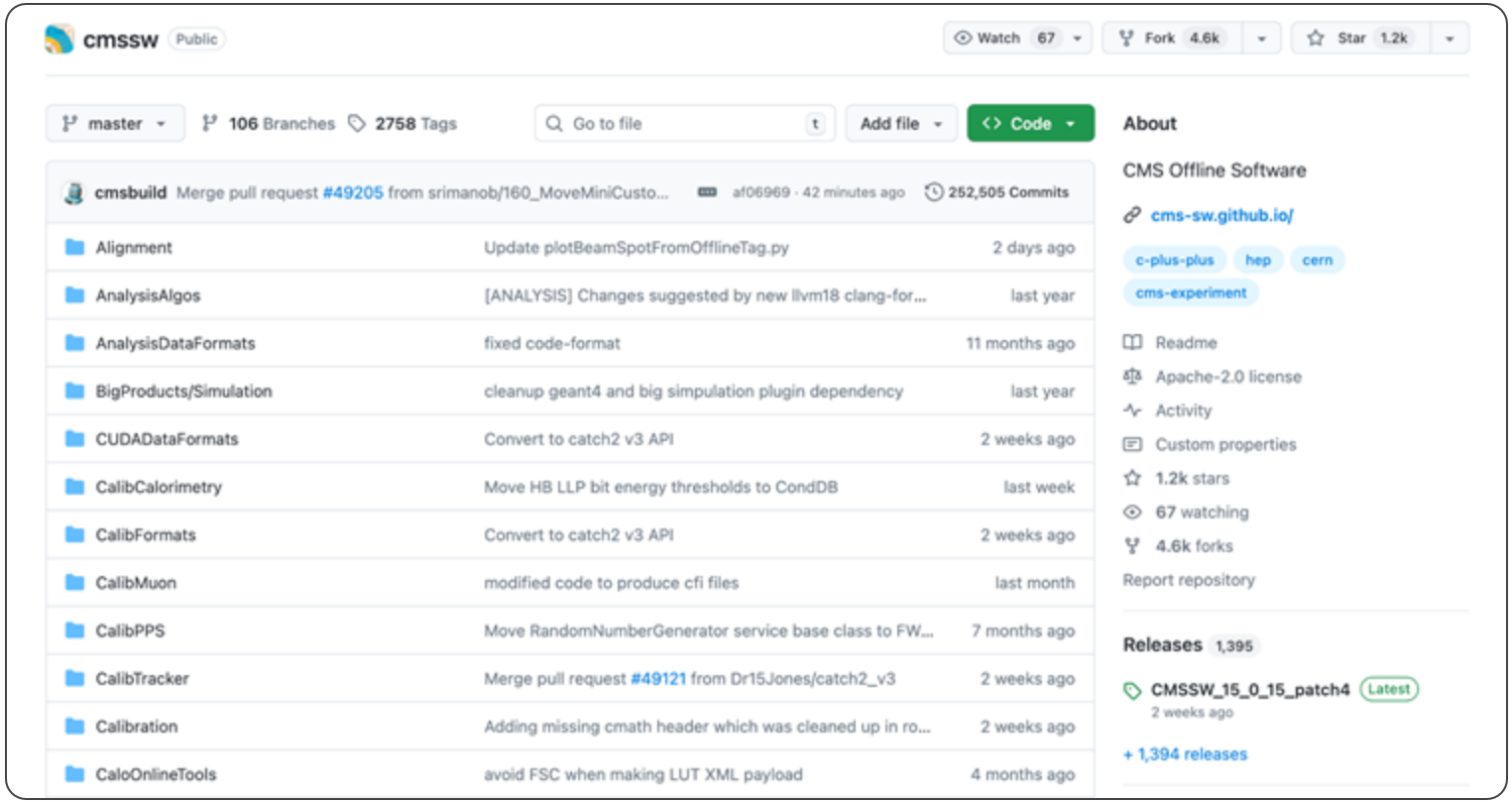
1. How do I run my analysis in a place that has **enough computing capacity** to process the data?

Containerised data analyses

https://www.reana.io

The software

https://cms-sw.github.io



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
  - Running workflows **without direct data access** (e.g. testing new models by theorists/phenomenologists by **reinterpreting** past analyses)
- 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