CMS experiment at CERN

Training opportunities in Data preservation and open access

08.07.2021
CERN - CMS DPOA - JUST

Hello!

I am Kati Lassila-Perini

- experimental particle physicist, PhD
- from Helsinki Institute of Physics (Finland)
- based at CERN
- coordinating data preservation and open access (DPOA) in the CMS experiment

kati.lassila-perini @ cern.ch

@KatiLassila

Hello from Clemens!

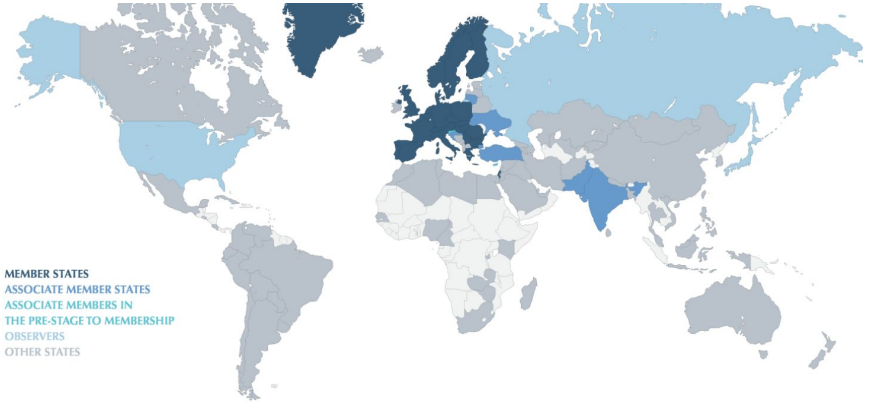
- Staff research physicist at CERN
- Leading and working in several different projects:
  - Searches for new particles
  - R&D for new particle detectors
  - Reusable physics analyses using cloud technology

1.

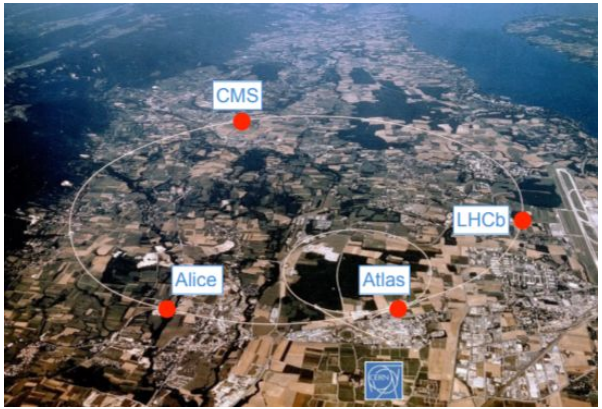
CERN?

European research organization that operates the largest particle physics laboratory in the world.

CERN: European particle physics laboratory



Large Hadron Collider LHC:



- Founded 1954
- 23 member states
- Host laboratory for experiments with participants from all over the world

2.

What is CMS?

One of the largest international scientific collaborations

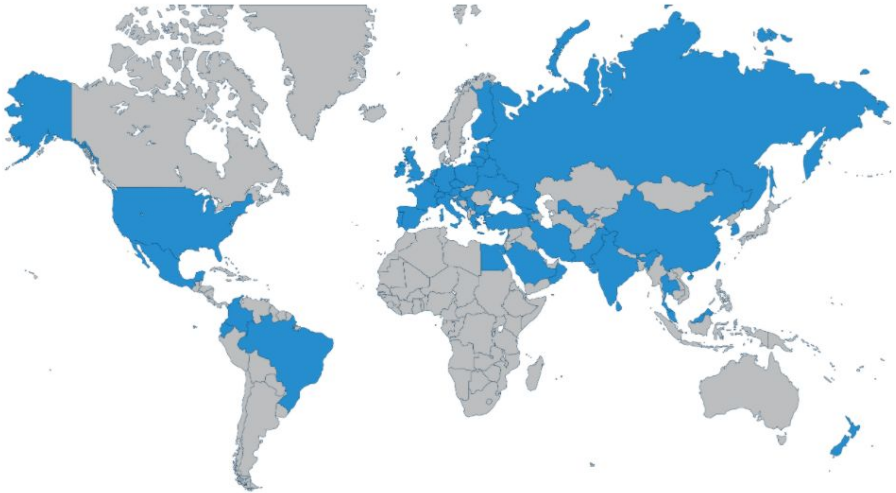
http://virtual-tours.web.cern.ch/vtours/CMS/CMS.html

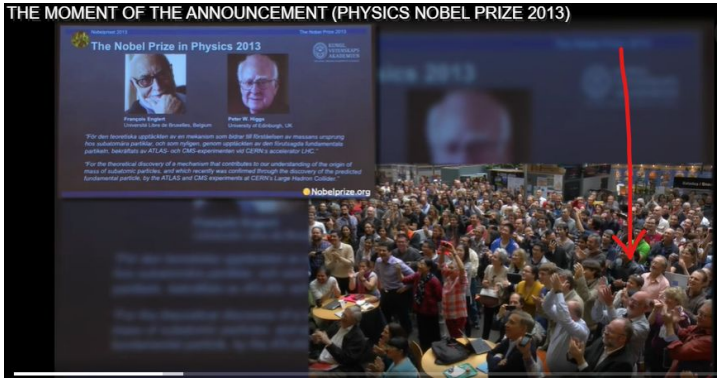
- **C**ompact **M**uon **S**olenoid
- Huge particle detector 100m underground
- Surrounds a beam collision point at the LHC
- Measures signals left by particles generated in the collisions
- Studies fundamental physics at highest available energies

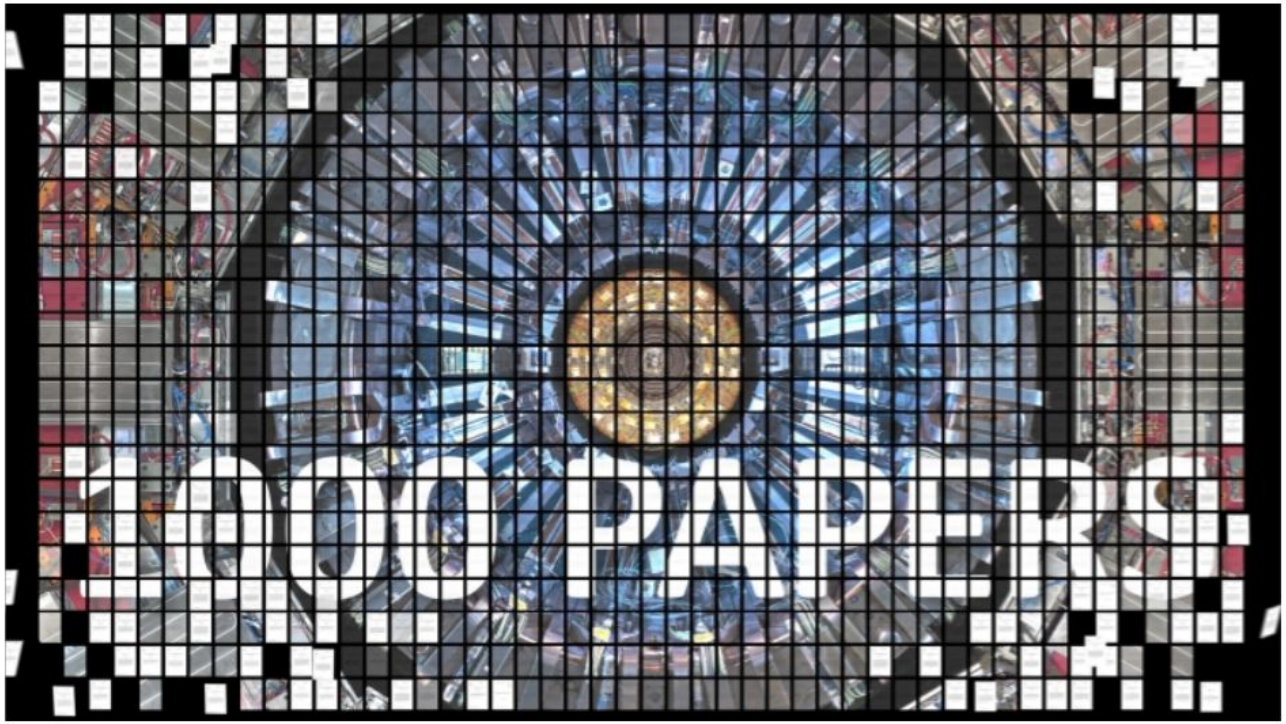
- Weight: 14 000 tons
- Diameter: 15m
- Length 28.8m
- Magnetic field: 3.8T

CMS collaboration:

2942	1065	281	229	51
PHYSICISTS (1036 STUDENTS)	ENGINEERS	TECHNICIANS	INSTITUTES	COUNTRIES & REGIONS







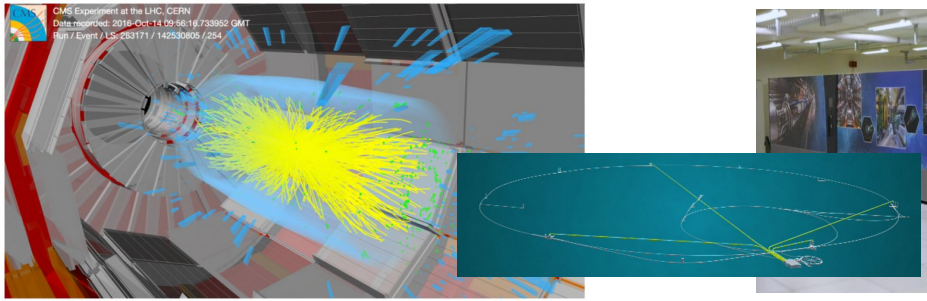
On Friday 19 June 2020, scientists at the CMS experiment at CERN's Large Hadron Collider submitted their 1,000th paper. This monumental achievement reflects an outstanding contribution to humanity's understanding of the universe — and it's just the beginning.

https://cms.cern/news/CMS-collaboration-celebrates-1000th-paper

3.
IT challenges

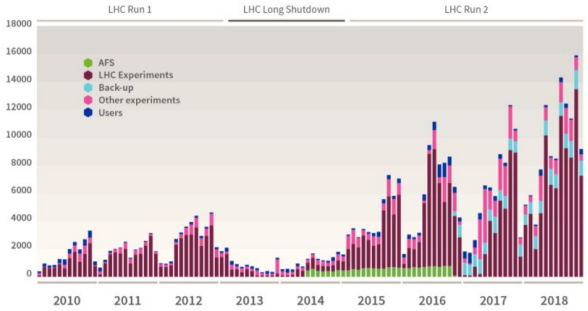
Huge amounts of data
Expensive, unique data

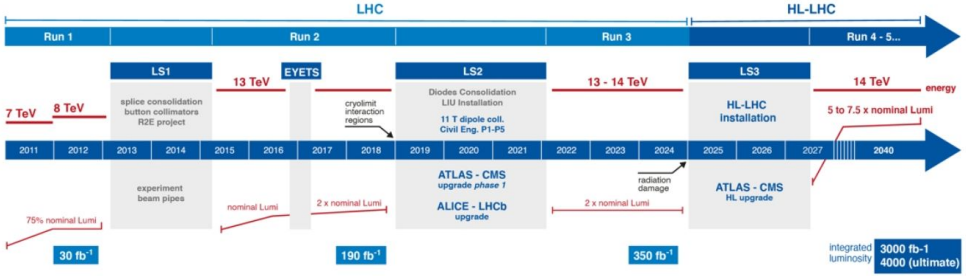
40 M collisions/sec:

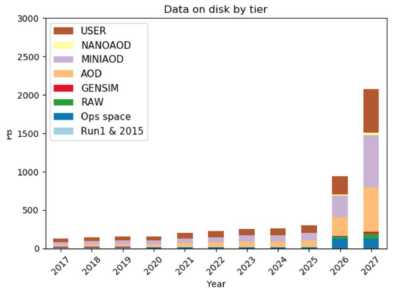


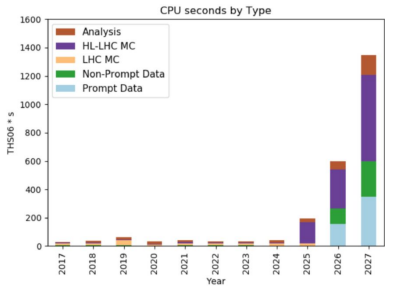
1000 collisions/s recorded:

TB/months recorded on tape:









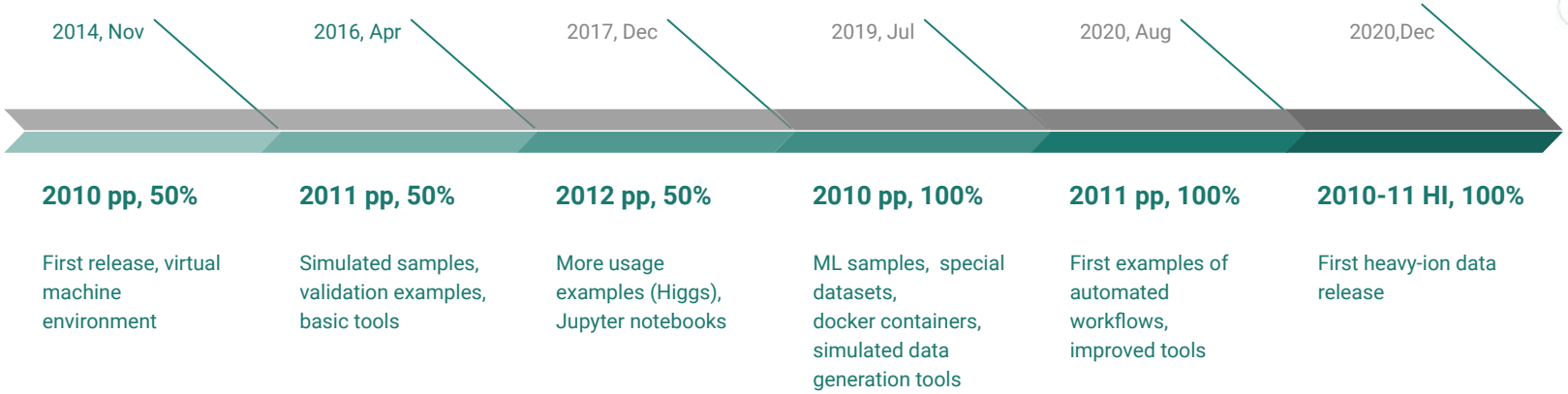
Reproducible analysis workflows

- Fact: **years of work** go in designing and implementing a physics analysis
- Goal: preserve analyses **during** the development/approval process already
  - Make this as **easy** as possible
- Challenge: find tools and methods **adapted** to physics analysis work
  - Helper tools for continuous integration, image building, …
  - Explore options for workflow automation
  - Training of physicists in these tools

Traineeship contributions needed in:

- Container image builds
- Using CERN-/CMS-specific technology in cloud systems
- Setting up common utilities

CMS open data



Traineeship contributions needed in:

- Preparation of next releases
- Collecting and displaying metadata

Efficient use of data: now and in the future

How to ensure the usability of data:

- Data and OS&SW needed for their use are old
- Find tools and methods in modern cloud environment

What is needed for a traineeship in CMS?

- Have an open mind
- Take a learning attitude
- Be prepared to finding out solutions yourself
- Share your skills
- Communicate
- Document

What do you get from the traineeship?

- Experience working in a truly international team
- Consolidate your skills with fundamental IT tools
  - CL, scripting, shell, git, unit tests, CI/CD, python…
- Gain knowledge of modern IT solutions
  - Workflow implementation, docker, k8s...
- Your learning is a value for us.

Thanks! شكرا

Questions?

Figures: CERN

Slide template: SlidesCarnival