

Open science in research and education: Project report



HIP Scientific Advisory Board - August 25, 2026

Kati Lassila-Perini, Peitsa Veteli
Helsinki Institute of Physics - Finland

The **Open Science in Research and Education** project (2026 - 2028) builds on the well-established foundations of the previous Open Data and Education project.

The education component continues and expands the successful science learning activities of previous years.

The research component extends the work previously carried out within the CMS experiment to a broader context in the high-energy physics domain, with a renewed emphasis on open science - as an outcome of open data and the adoption of best practices in software development and computing throughout the data lifecycle.



“

1

Open science in research

Data preservation and open data in the CMS experiment
The ICFA Data Lifecycle panel



CMS Open data

- Under the 12 years of HIP leadership (2012 – 2024), the CMS open data releases were built into a mature, policy-driven, sustainable pipeline.
- The release work now continues within the small, dedicated open data team in CMS.
- Recent HIP efforts towards long-planned technical extensions
 - productive team work by a group of summer trainees (2 from HIP, 2 from Jordan) during this summer
 - benchmarking time and resources needed for typical open data processing end user tasks
 - setting up internal, continuous testing for reprocessing open data from raw data, or from simulation input.

CMS Open Data on Infomaniak

Setup

Episodes

Resources

Versions

Search...

CTRL K

CMS open data processing on Infomaniak cloud resources

Learn to run a typical CMS open data processing job using public cloud resources from Infomaniak.

Get Started

Setup

Prepare your computing environment to interact with computing and storage resources on cloud.

Learn more

Learn more about CMS open data by working through the tutorials in one of our workshops.

Understand pricing

This tutorial gives an estimate of the cost and time needed for a typical processing task

Simulation Workflow Tutorial

Setup

Episodes

Summaries

Versions

Search...

CTRL K

Simulate your own collision dataset using Infomaniak cloud resources

Learn to run a CMS data processing workflow and simulate your own dataset using public cloud resources.

Start here

Setup your environment

Install Kubect!, Argo, the workflow code and other essentials

Kubernetes documentation

Get to know more about cluster architecture on the Kubernetes website

First run

After setting up your environment, do the first run of the workflow

README

GPL-3.0 license

ReReco REANA Snakemake Workflow

A REANA workflow for processing CMS RAW heavy-ion data into RECO format and producing HiForest ntuples with PF candidate plots.

- reco_base: Just raw2reco step using container built on top of base slc5.
- reco_standalone: Just raw2reco step using standalone image.

Pipeline

RAW (EOS) → raw2reco (CHSSW 5.3.8) → reco2forest (CHSSW 5.3.20) → plotforest → plots

Step	Rule	Output
RAW → RECO	raw2reco	res/pahightp_reco.root
RECO → HiForest	reco2forest	res/HiForest.root
HiForest → plots	plotforest	res/plots/?.png

Input data: CMS Open Data HIRun2013 PAHighPt RAW Outputs: res/HiForest.root and four PF candidate plots (pT, η , ϕ , η - ϕ)

Reana Simulation Workflow Tutorial

Setup

Episodes

Materials

Versions

Search...

CTRL K

Workflow Repository

Simulate your own collision dataset using REANA and Snakemake

Learn to run a CMS data processing workflow and simulate your own dataset using CERN's REANA platform.

Get Started

5



ICFA Data Lifecycle panel

- KLP was called to lead a panel on the Data Lifecycle
 - a panel of the International Committee for Future Accelerators (ICFA)
- The panel focuses on Open science and FAIR practices
 - FAIR: Findable, Accessible, Interoperable, and Reusable digital assets (data, software, workflows, ...)
- As a first major action, panel has formed best-practice recommendations for data preservation and open science in HEP

Recommendations for best practices for data preservation and open science in HEP

Home Foreword Executive summaries ▾ Graphs ▾ More ▾ v1.1

☐ host laboratory ☐ experiment management ☐ home institute ☐ WG leaders ☐ funding agency ☐ tool developers ☐ analysts ☐ data management ☐ open data group

Filter by Class

☐ Policy and management ☐ Infrastructure and services ☐ Software skills development
☐ Licenses, copyright and citations ☐ Community-wide software development
☐ Collaboration-specific software development ☐ Software and workflow management - analysis-specific SW
☐ Analysis preservation tools and practices ☐ Data management tools and practices
☐ Documentation and knowledge preservation ☐ Long-term sustainability ☐ Cost, funding and return of investment
☐ International collaboration

ID	Class	Recommendation	Actors	Depends on	Enables	Description
PM1	Policy and management	Develop a comprehensive archival policy that encompasses all relevant scientific research outputs generated through the organization's research activities, supported by clear governance	host laboratory		PM13, PM9, PM3, PM17, PM16, PM14, PM19	Show description

<https://icfa-data-best-practices.app.cern.ch/>
<https://arxiv.org/abs/2508.18892>
[ICFA statement](#)



From recommendations to action

Organizing a domain-wide evaluation

Request assessment teams from

- **Host laboratories**
- **Experiment managements**

to assess how the recommendations are implemented in their case.

Recommendations for best practices for data preservation and open science in HEP v1.1

Home Foreword Executive summaries Graphs More

Not connected

Username: Password:

Select...
exp_alice_2026
exp_amber_2026
exp_atlas_2026
exp_bes3_2026
exp_cms_2026
exp_daya_2026
exp_example_2026
exp_juno_2026
exp_lhaa_2026
exp_lhcb_2026
exp_na62_2026
exp_na64_2026
exp_ship_2026
exp_snd_2026

Assessment score 13%

preservation and open science in HEP

			publication.			
APS	-- Select status -- -- Select status -- Applied Partially applied Planned Not yet considered Not applicable	Analysis preservation tools and practices	Require that all code and configuration necessary to produce the final plots in an analysis be findable, available, and reusable as a condition for approval.	experiment management, WG leaders	APT	APS API2 PM2

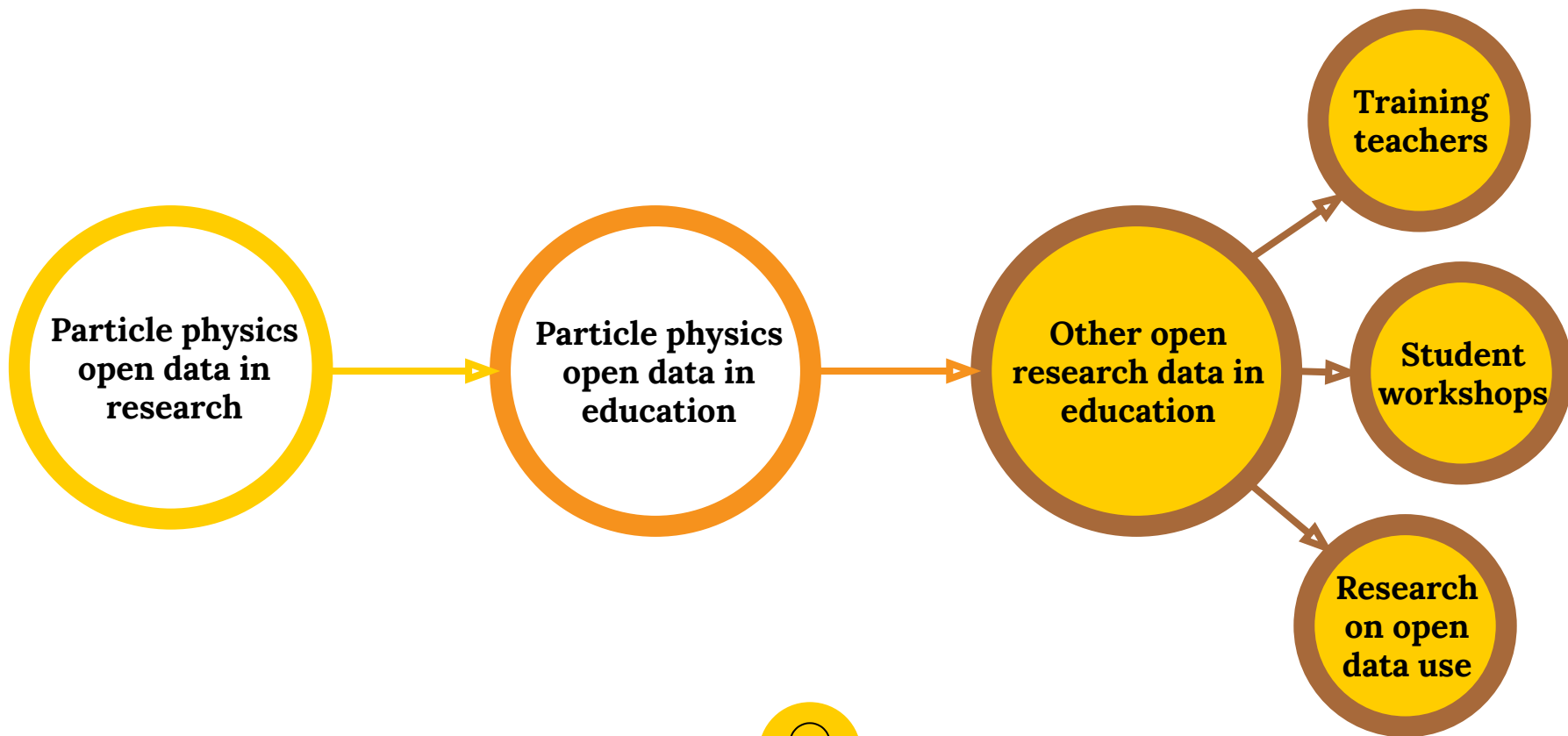
Assess

Report

2

Open science in education

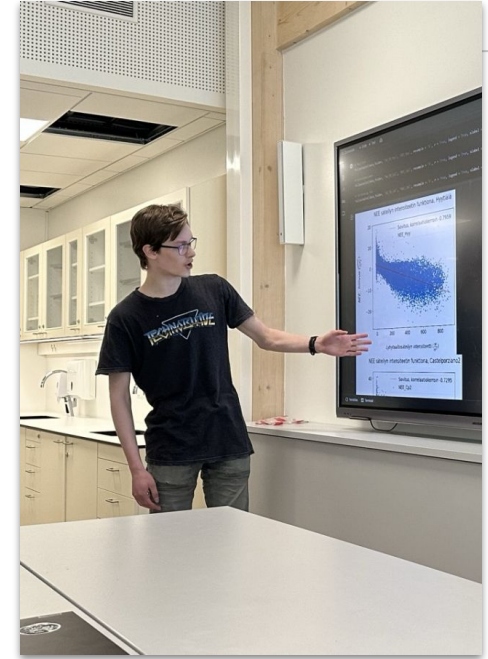
Peitsa Veteli, doctoral researcher





Example activities 2025-2026

- Teacher trainings x 2 (~ 30 participants)
- Student workshops x 12 (~300 students)
 - combined to a specific course (e.g. a six week project course at HELLU) or as an independent activity
 - climate, space physics, particle physics, aerosols, general science, data science, math, life sciences, social studies, health...



Example activities 2025-2026

- ◉ Introductions to open data activities
- ◉ Participation in in-service teacher's curricular planning and material production
- ◉ Pre-service teacher courses
- ◉ Research presentations
 - ◉ GIREPs, Finnish stem-days, subject didactics...
 - ◉ Authentic data in teaching
 - ◉ Developing the physics E/O model at Helsinki





Upcoming events in OS&E

- 25.8. STEM-teacher training alongside geography and chemistry departments' edu-teams (~30 teachers)
- 25.9. Researchers' Night (~100 lower secondary registrations)
- 14.11. Harppi-fest (workshops for upper secondary students)
- xx.11. Authenticity in Science Education – pre-service teacher course at Hyytiälä station (~10 students)
- xx.1.2027 Science Days data workshops (upper 2nd stud.)
- Ongoing development of the Fotoni lab as an E/O hub together with HIP, INAR and the Department of Physics
- ...



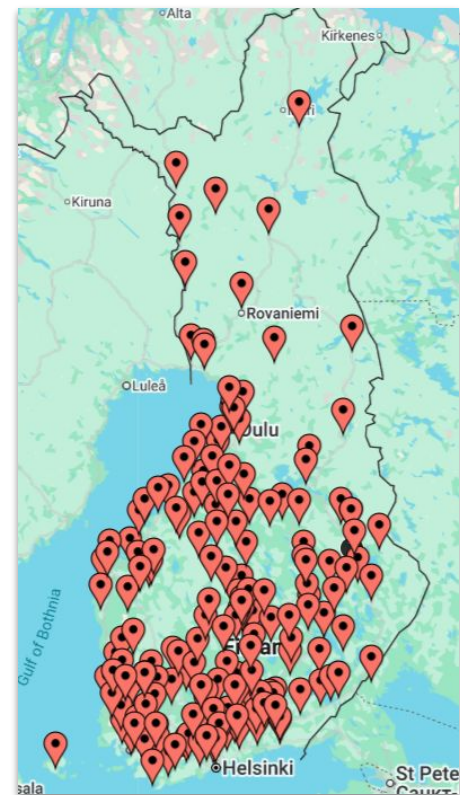
3

High school visits to CERN



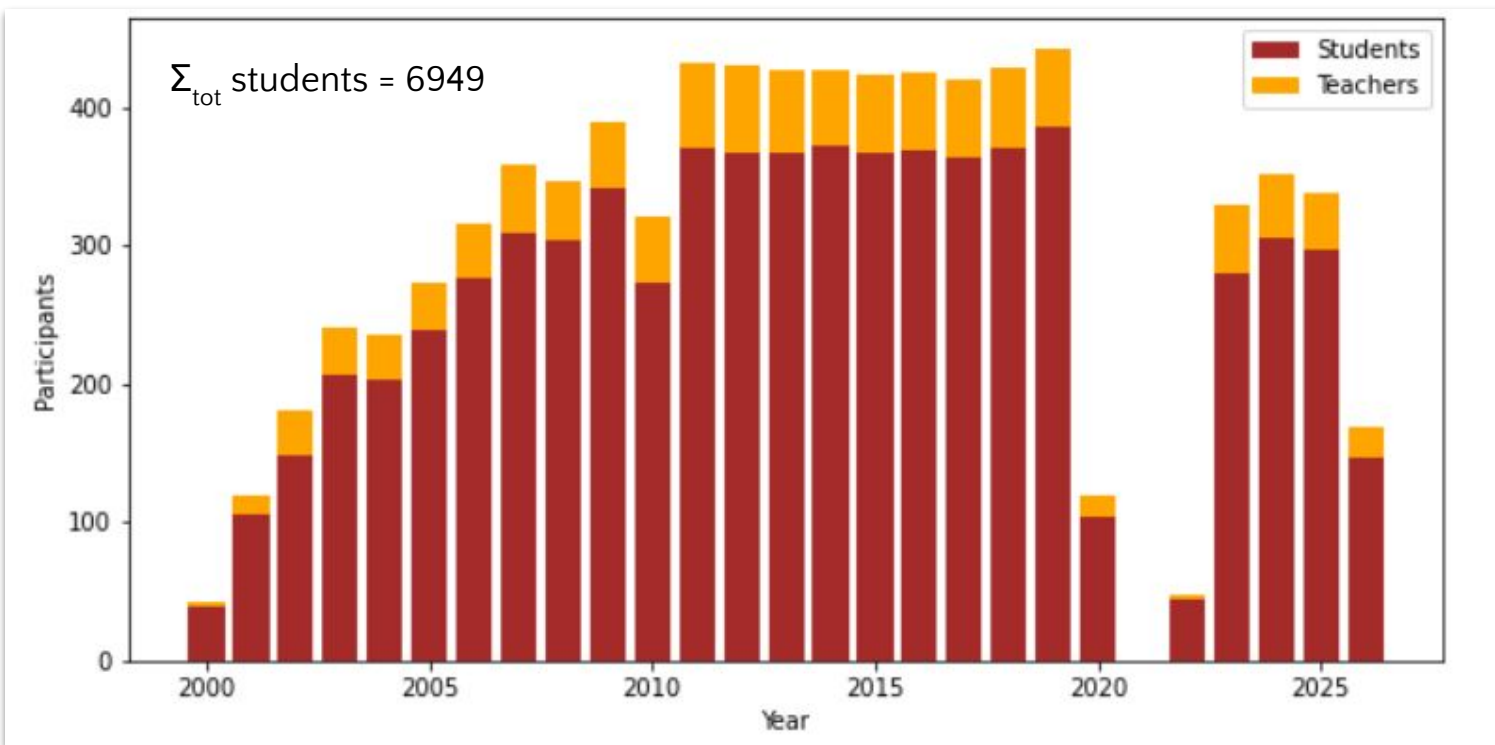
High schools visits 2000 -2026

- Schools from all Finland
 - Majority of Finnish high schools have had visits to CERN
 - Every group has participants from several schools
- More interest than possibility to organize visits
 - National coordinators select the groups based on applications
 - Organization at CERN by the HIP team
- Funding
 - Previously by Finnish National Agency for Education
 - Currently by the Ministry of Education



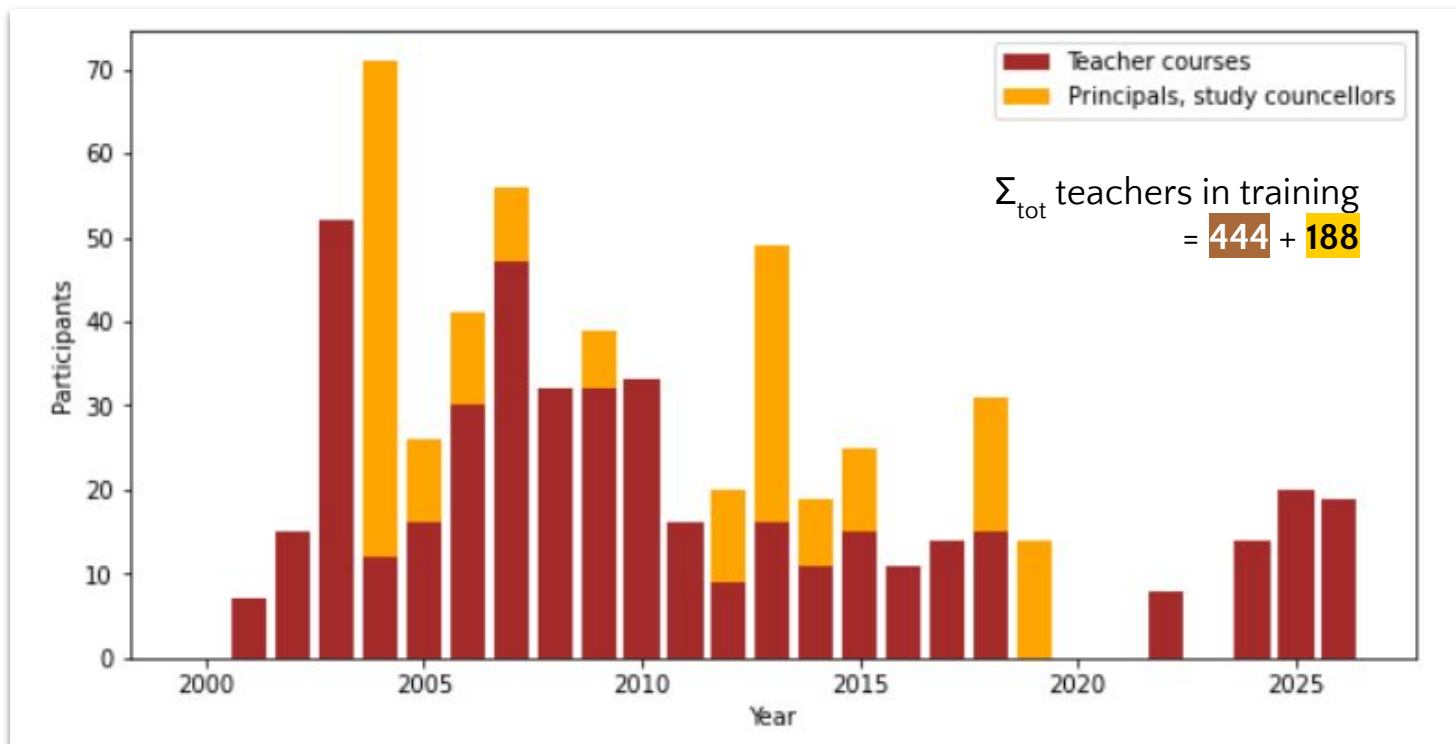


Each year 300-400 Finnish students visit CERN





Training for Finnish teachers at CERN





The well-established and popular high-school visits to CERN co-coordinated by the national organizers and HIP.

