

# Paving the Way for Open Data

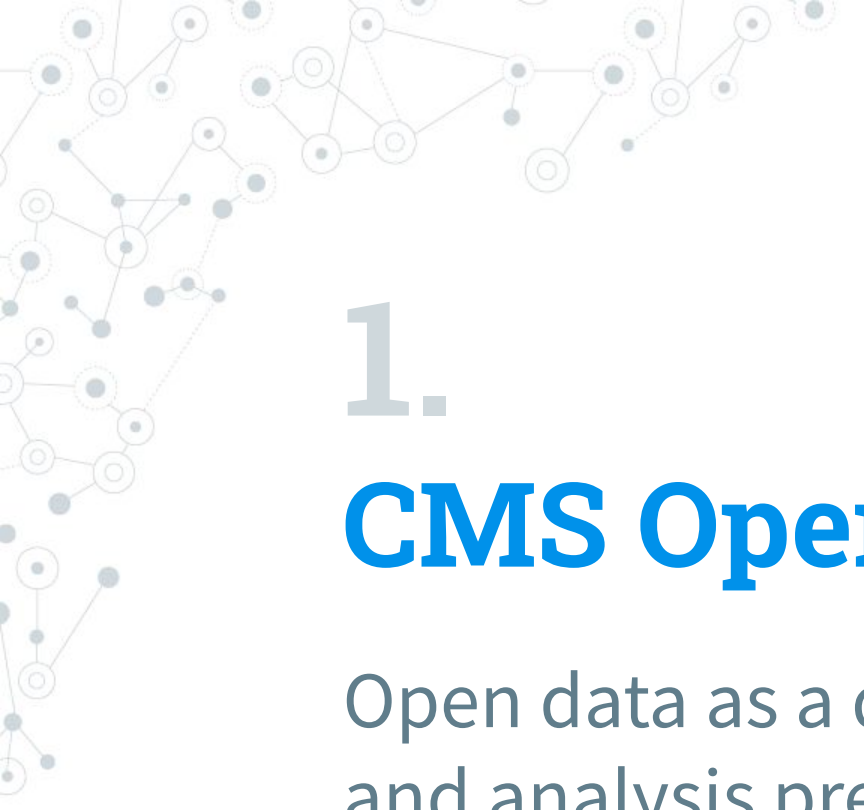
## 10 years of releases from CMS

**Julie Hogan, Bethel University**

On behalf of the CMS Collaboration

CMS Data Preservation and Open Access group

July 11, 2025

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue and others in grey.

1.

# **CMS Open data - Why?**

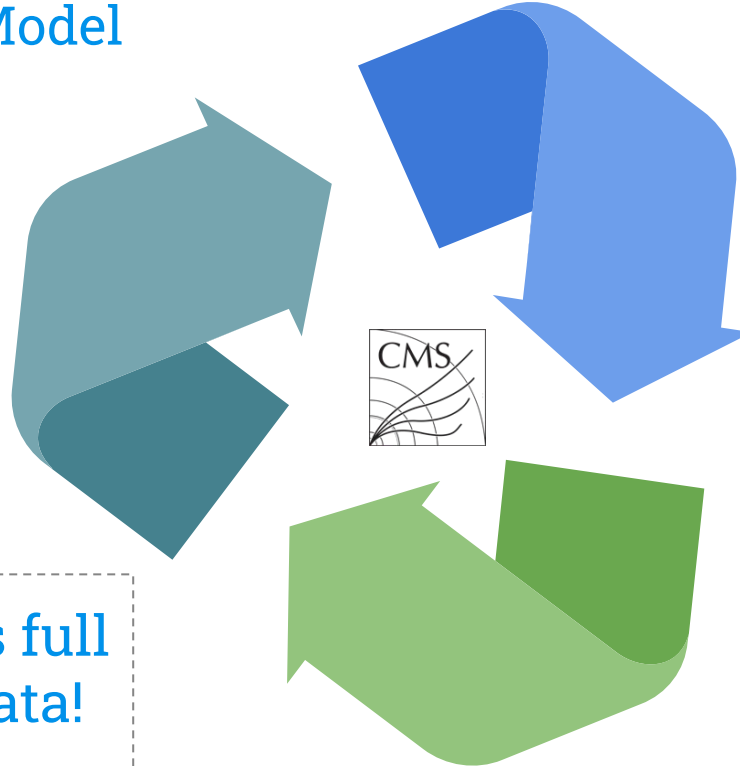
Open data as a driving force to data  
and analysis preservation

# CMS Open Data Model

## Data:

- collision data
- simulations
- additional data for analysis

CMS Open Data is full  
research-level data!



## Tools:

- software
- environments
- interfaces

## Knowledge:

- instructions
- actionable examples
- understanding of experimental data

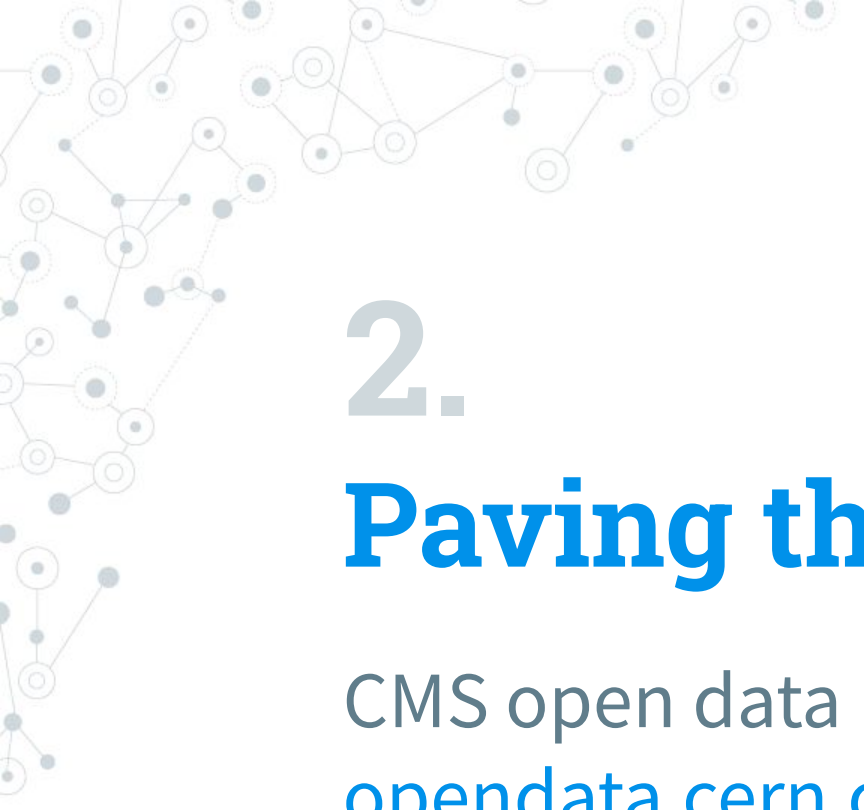


Matthew Strassler, Jesse Thaler  
Nature, August 1, 2019  
note to the editor

“

But steady publication of LHC data has multiple benefits. First, it encourages prompt archiving, before collective memory fades and knowledge is lost. Second, other scientists can analyse the data while the LHC is still running, testing unconventional strategies and potentially leading to unexpected discoveries, new approaches and fruitful discussions. And third, as a by-product, these scientists can stress test the archiving methods; any deficiencies found are easier to fix now than later. In this way, public collider data can complement the overall LHC research effort. We, therefore, favour a slow but steady approach to full publication of the LHC experiments' data; it is in the best interest of particle physics.

Open data have  
value only when  
they are in use!

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

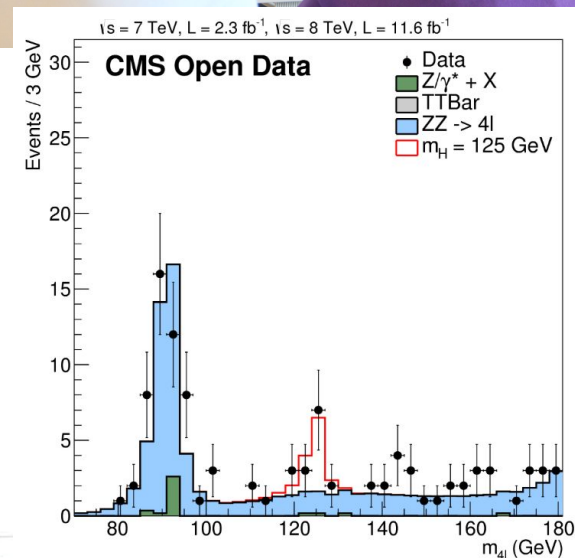
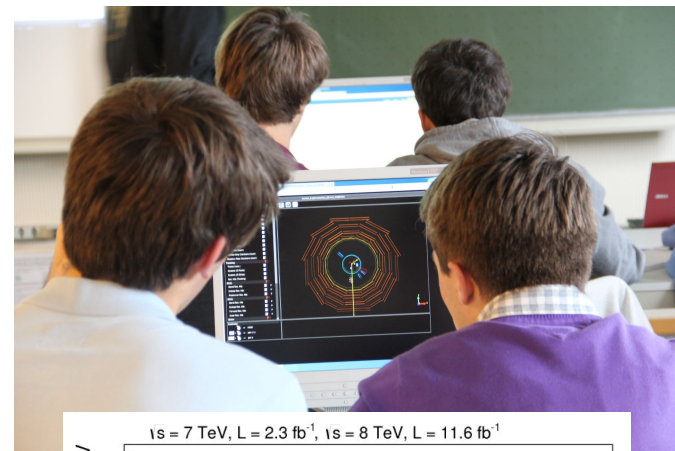
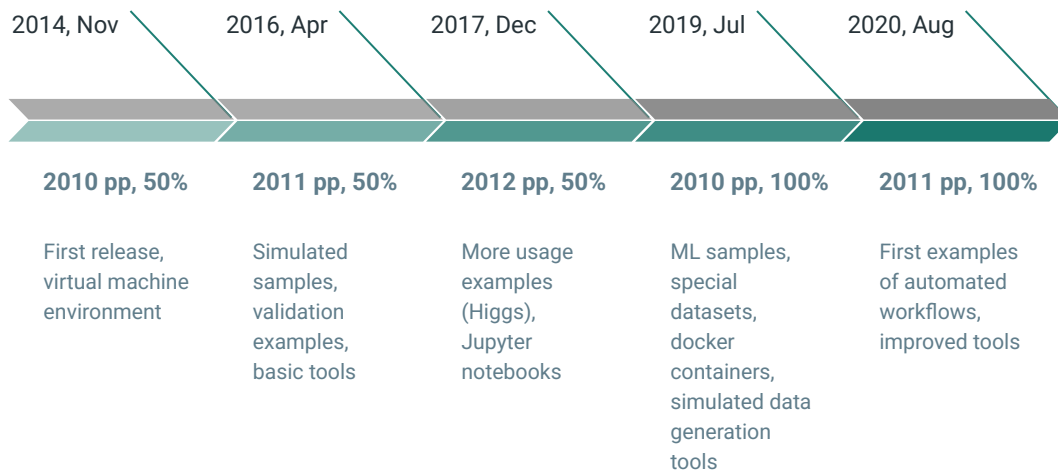
# 2.

## Paving the way

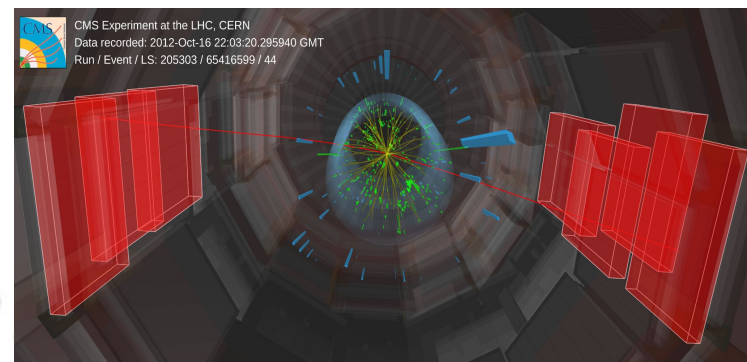
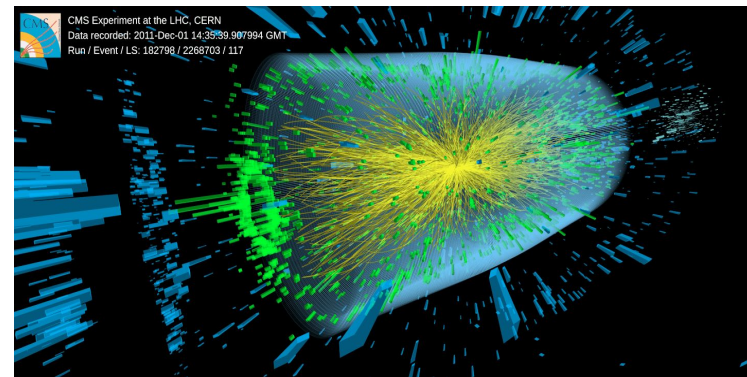
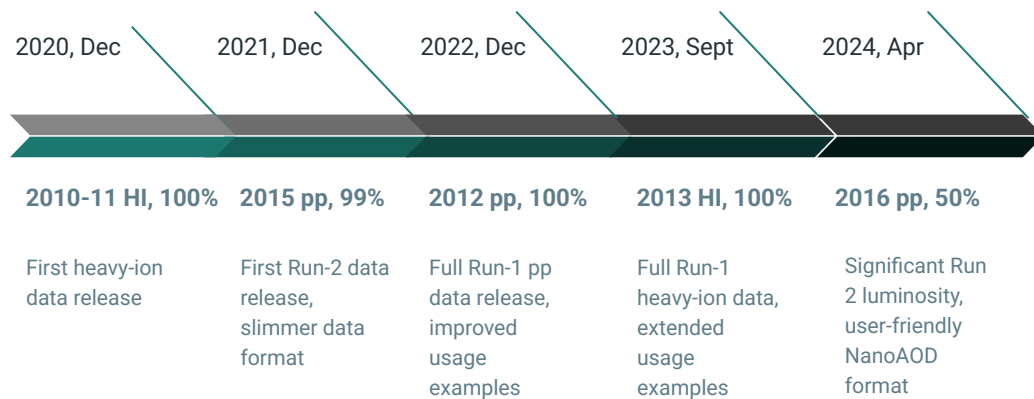
CMS open data releases since 2014

[opendata.cern.ch](https://opendata.cern.ch)

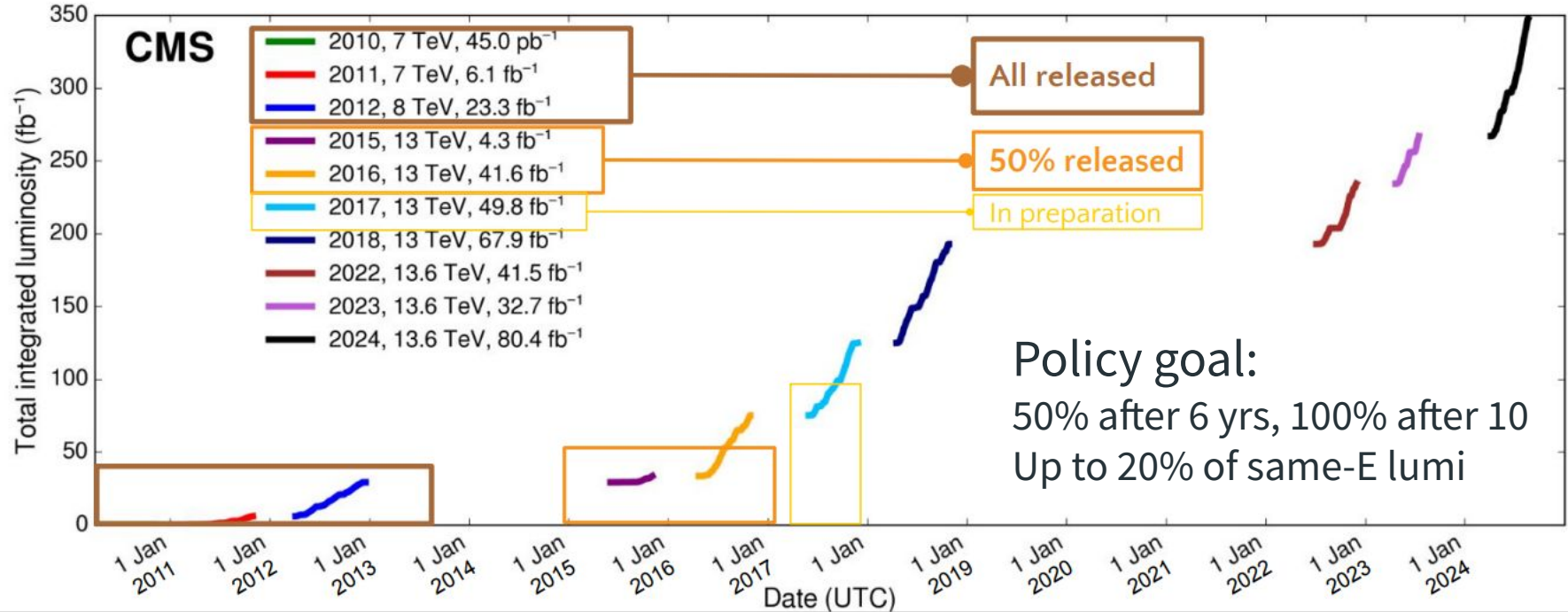
# Release timeline



# Release timeline

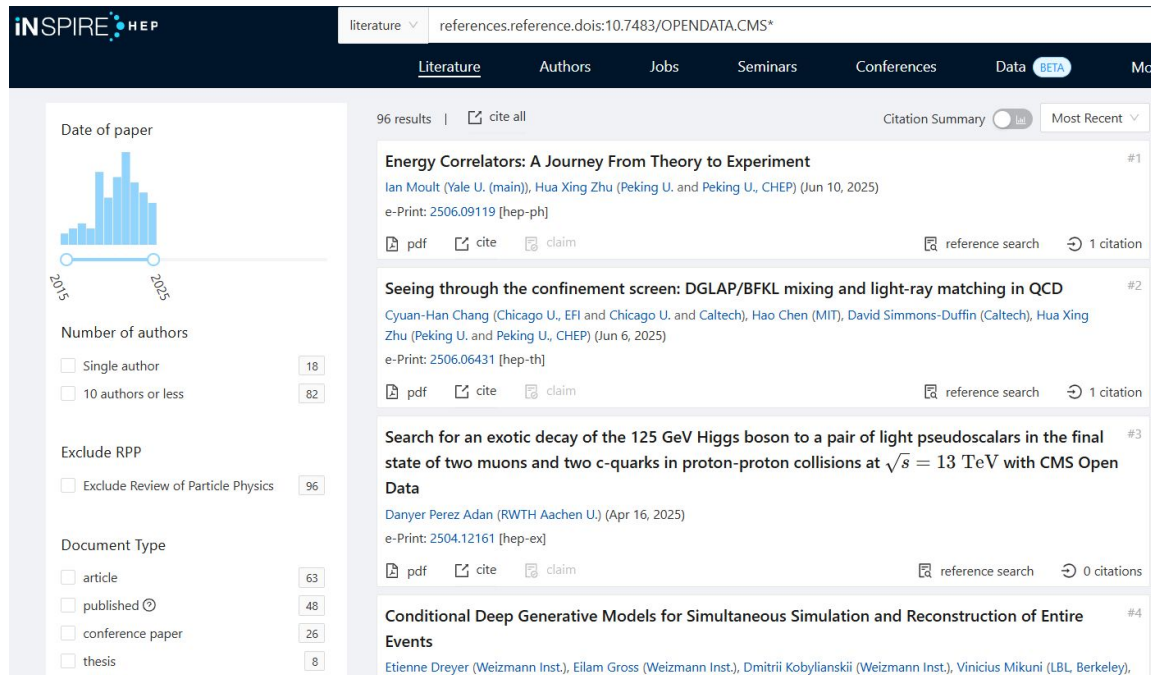


# CMS Open Data releases

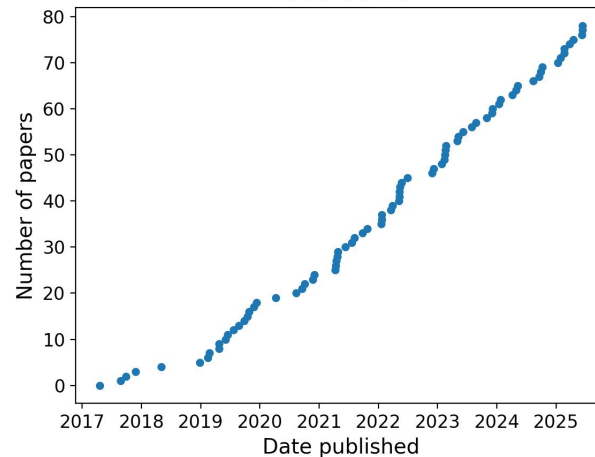


# CMS Open data in use

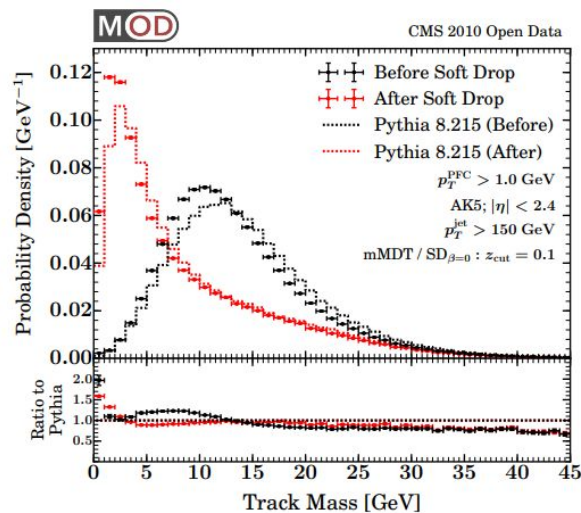
Records have unique DOIs: [10.7483/OPENDATA.CMS\\*](https://doi.org/10.7483/OPENDATA.CMS*)



Papers citing CMS Open Data DOIs [Inspire]  
2025-06-30



# Publication examples 2017 – 2025

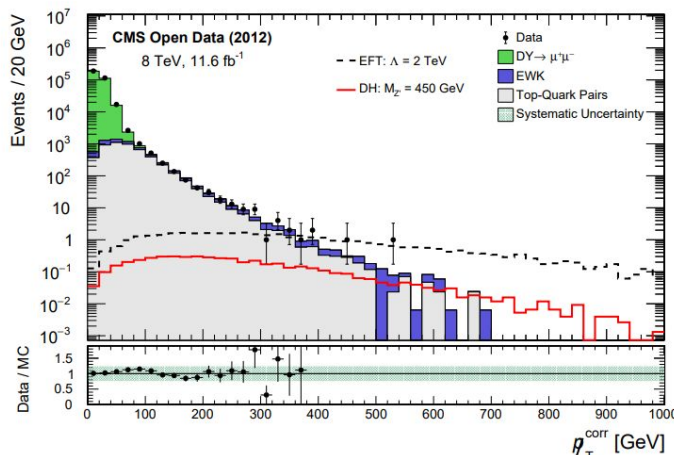


Theoretical QCD studies

e.g. [1704.05066](#)

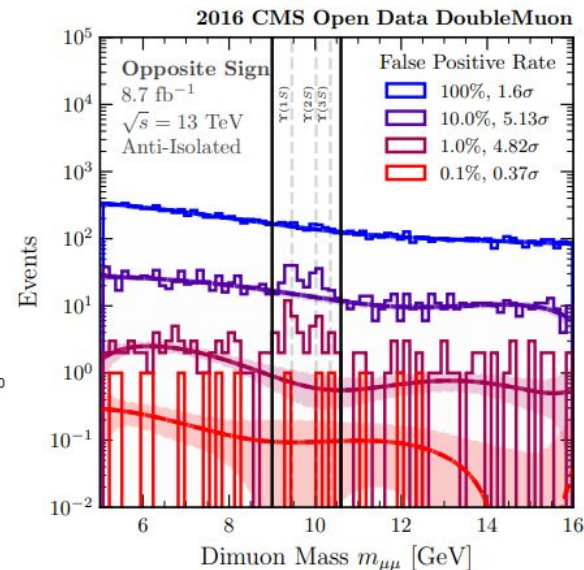
New physics searches

e.g. [2109.11274](#)



Anomaly detection for searches

e.g. [2502.14036](#)



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue and others in grey.

# 3. New Opportunities

Run 2 data



CMS Experiment at the LHC, CERN

Data recorded: 2016-Aug-15 01:00:30.361728 GMT

Run / Event / LS: 278822 / 837399836 / 484



**Read more!**

**CMS releases 13 TeV proton collision data from 2016**

# 2016 data release features

## Collision Data

- ◎ 16 fb<sup>-1</sup> of 13 TeV proton collision data from 2016
- ◎ Ultra-Legacy processing!
- ◎ MiniAOD and NanoAOD data formats

## Simulation

- ◎ Broad array of SM & BSM simulation
- ◎ Over 20,000 unique processes!
- ◎ MiniAOD and NanoAOD formats

## Software

- ◎ Container & VM for CMSSW 10
- ◎ Containers for ROOT & python
- ◎ New guides
- ◎ New analysis tools

First significant luminosity from 13 TeV collisions!

# NanoAOD

**95%  
smaller!  
1-2 kb/evt**

**Flat ROOT  
TTree  
Basic types**

**Particle Flow  
info added  
for some data**

**Extract data  
straight  
into arrays!**

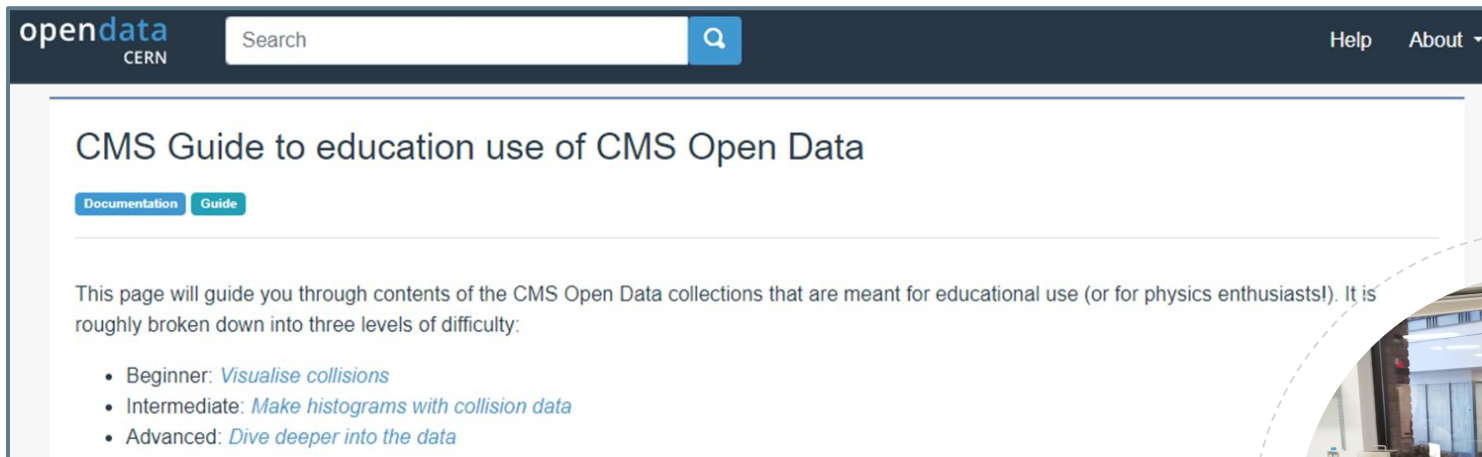
**A big step toward easily reusable CMS data**

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric rings, suggesting a hierarchical or multi-layered structure. The lines are thin and gray, connecting the nodes in a non-linear fashion.

# 4. **Resources & Future goals**

# Education & Outreach

<https://opendata.cern.ch/docs/cms-guide-for-education>



The screenshot shows the 'opendata CERN' website header with a search bar and 'Help'/'About' links. The main heading is 'CMS Guide to education use of CMS Open Data'. Below it are tabs for 'Documentation' and 'Guide'. The text states: 'This page will guide you through contents of the CMS Open Data collections that are meant for educational use (or for physics enthusiasts!). It is roughly broken down into three levels of difficulty:'

- Beginner: *Visualise collisions*
- Intermediate: *Make histograms with collision data*
- Advanced: *Dive deeper into the data*

- Interactive & VR event displays
- International Masterclass
- Dimuon analyses for schools
- University-level course tools



# Research Use

<https://opendata.cern.ch/docs/cms-guide-for-research>

## CMS Guide to research use of CMS Open Data

Documentation

Guide

If you are interested in step-by-step instructions to start working with CMS Open Data, please consult these pages:

- [Install Virtual Machine](#) or [Use a container](#)
- Getting started with CMS [AOD Data](#), for data collected during Run 1 of the LHC.
- Getting started with CMS [MiniAOD Data](#) or [NanoAOD Data](#), for data collected during Run 2 of the LHC.
- Getting started with CMS [Heavy Ion Data](#).

This page offers hints, tips and guidance for conducting a research-oriented analysis using CMS Open Data. More detailed information can be found in the [CMS Open Data Guide](#).

- Logistics & Software  
*[“Getting Started” Guides](#)*  
*[Containers & VMs](#)*  
*[Metadata & Artifacts](#)*
- Physics analysis  
*[CMS Open Data Guide](#)*  
*[Example analysis records](#)*
- Hands-on help  
*[Annual Workshops!](#)*

## Coming Next

- ◎ Partial release of 2017 data coming in 2025
  - ◎ Continued participation in Open Science discussions across HEP
  - ◎ **Goal:** complete Run 2 releases by 2030, alongside first Run 3
- 
- ◎ Regular updates of the Open Data Guide for Run 2 data
  - ◎ Run 2 NanoAOD analysis examples
  - ◎ Heavy Ion analysis examples
  - ◎ **Goal:** mini immersion sessions at conferences

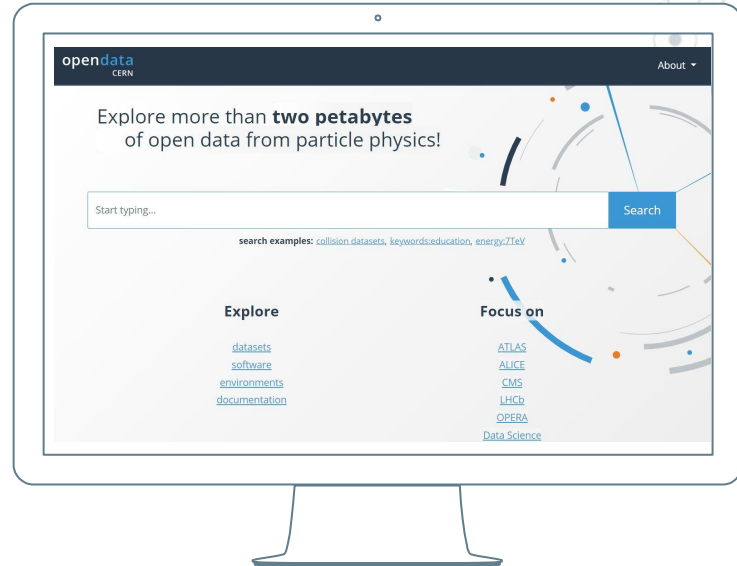
Thank you!

Thanks to our colleagues:

- ◎ CMS DPOA group
- ◎ CERN data preservation team
- ◎ CERN Open Data Portal team

And great thanks to all CMS open data users!

Template by [SlidesCarnival](#)



## CERN Open Data portal

Come explore what **you** can do with CMS Open Data!