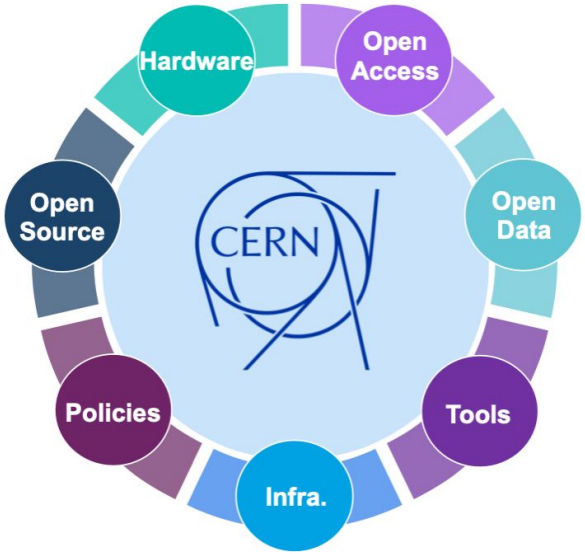


Open Data and AI

Towards a National-Scale AI Collaboration in HEP

18 May 2026

Zach Marshall (LBNL)

BERKELEY LAB

The Levels of Open Data from DPHEP

Level 1: Additional Documentation

We do this all the time through plot records, HepData, Rivet Analyses…
Almost all physics papers should have some combination of these!

Level 2: Simplified Formats for Education and Outreach

CMS and ATLAS have quite a bit (10 year anniversary for ATLAS Open Data!).

Level 3: Analysis Formats and Software

Includes both analysis formats and custom datasets for a wide variety of applications.

Level 4: Experiment Raw Data

Will preserve it; won't make it public at large scale (there's too much and it's too hard to document).

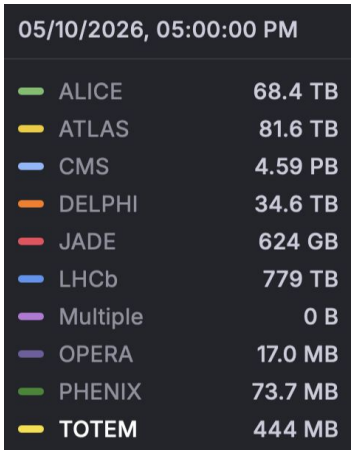
Bespoke datasets

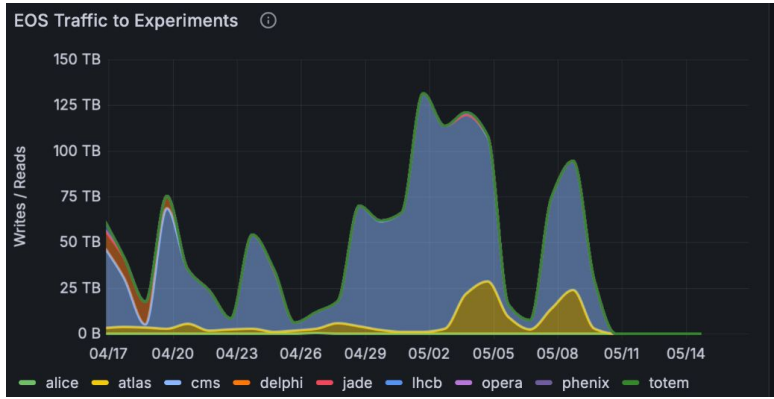
- We have lots of **bespoke datasets** for **targeted research applications**
  - ATLAS's JetSet, Top jet tagging (w/systematics), Fast Sim training, BSM BDT training, …
    - Within ATLAS we try to always have an example git repo with the dataset
  - Participation in Kaggle Challenges: Higgs boson ML Challenge, TrackML Challenge, …
  - LHCb's simulated jet samples for quark flavour ID studies (used for a quantum ML paper)
  - CMS's ML datasets with example workflows
  - CMS recently released some **RAW data** for the CICADA project. Interesting test: minimal documentation (that I can find), but CMS members allowed to use internal documentation.
- Notice the common theme here: machine learning (more discussion here)
  - Seems to be the 'best' motivation for specialized research data
- There is also sharing of **files** for technical purposes (e.g. ROOT/compression tests)
- All experiments seem amenable to adding new datasets with a well-motivated request

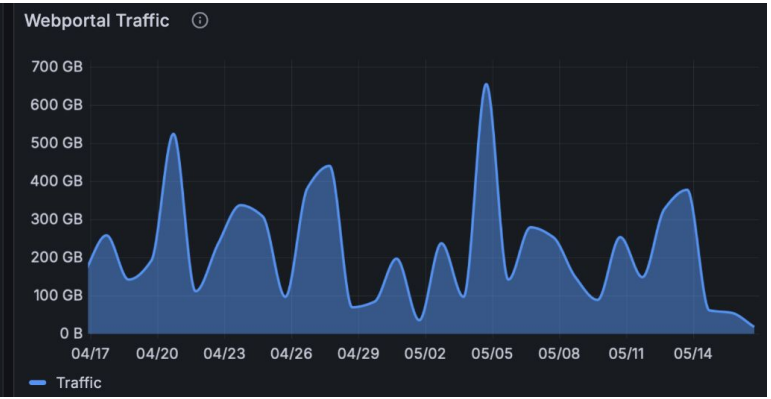
ATLAS folks can see upcoming / proposed releases here; quite a few in flight

Big Open LHC Data Sets

- We have released substantial Open Data via the CERN Open Data Portal
  - ATLAS's total is closer to 1PB, the monitoring is just a bit behind
- These include >10^{10} events, ~10^{5} datasets, and several different *kinds* of data
  - pp and heavy ion collisions, mostly Run 1 and 2015–2016 of Run 2
  - ROOT-based "light" data formats (nanoAOD/PHYSLITE), heavier formats (mini/AOD), special formats (HEPMC)
  - Also educational versions (more about that in a bit): simplified, flat ntuples
- All the data are released under a CC0 license (free use) with a request for citation
- Usage policies vary significantly across experiments







Big Open Neutrino Data Sets

Table from this talk by Jianming Bian

Experiment	(1) Data	(2) Metadata	(3) Analysis Data	(4) Tools
MicroBooNE	2D wire-time images, event data — HDF5, artroot — Zenodo	dataset docs — Fermilab site	reduced datasets — Zenodo	Python data loaders, example Jupyter notebooks — GitHub (OpenSamples)
MINERvA	tabular ntuples — ROOT — Fermilab server	docs — Fermilab site	reco vars + systematics — Fermilab server	ROOT macros (event reading, plotting), systematics weighting scripts — GitHub (MinervaExpt)
Daya Bay	tabular event data — HDF5, NPZ, ROOT — Zenodo	metadata — Zenodo	analysis dataset (IBD + inputs) — Zenodo	Python analysis package (dayabay-model), fit scripts — GitHub
NOvA	histograms, tables — ROOT/plots — Fermilab publicdocs	docs — Fermilab site	oscillation / xsec products — Fermilab publicdocs	
T2K	fit outputs, histograms — ROOT — Zenodo	metadata — Zenodo	fit results — Zenodo	ROOT macros (plot extraction, validation) — Zenodo
IceCube	event-level data (time, charge, position) — HDF5/CSV — IceCube data release site	docs — IceCube site	event catalogs — IceCube site	
PILArNet	2D/3D sparse images, point clouds — numpy/HDF5 — OSF	dataset paper — arXiv	labeled dataset — OSF	dataset interface scripts (data loading, preprocessing) — project repo
NuBench	detector hits (graph/point-cloud, tabular) — Parquet, SQLite — ERDA	docs — arXiv + GitHub	benchmark tasks — GitHub	data loaders, training scripts, model implementations (GraphNeT-based) — GitHub

Tooling

- ATLAS implemented atlasopenmagic to provide metadata (tutorial)
- Other experiments currently using the CERN Open Data Portal (client)
- MCP servers are now being set up (atlasopenmagic, ami, rucio, more…)
- Lumi is the most ‘advanced’ assistant I’m aware of in this space
  - Skills including CERN Open Data Portal; atlasopenmagic; rucio; hepdata; publication reading from CDS, arXiv, etc; ReAna; FTS; PDG; mkdocs CERN pages
- Lots of folks working with Agentic AI; Open Data is a fertile training ground

Automating High Energy Physics Data Analysis with LLM-Powered Agents

Eli Gendreau-Distler,[1,2,*] Joshua Ho,[1,2,†] Dongwon Kim,[1,2,‡] Luc Tomas Le Pottier,[1,2,§] Haichen Wang,[1,2,¶] and Chengxi Yang[1,2,**]

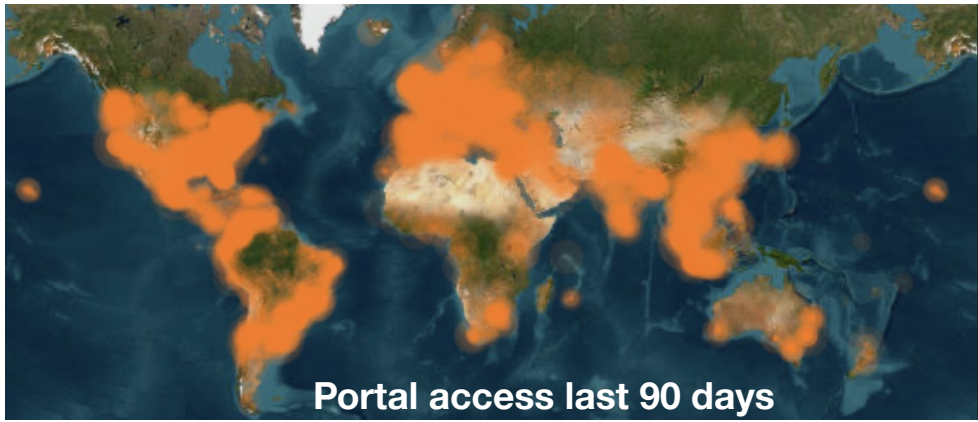
[1] *Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA*
[2] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

We present a proof-of-principle study demonstrating the use of large language model (LLM) agents to automate a representative high energy physics (HEP) analysis. Using the Higgs boson diphoton cross-section measurement as a case study with ATLAS Open Data, we design a hybrid system that combines an LLM-based supervisor–coder agent with the **Snakemake** workflow manager.

Access and Resource Monitoring

- Significant hardware resource limitations at CERN
  - Some open data being moved to tape, with a system for staging on (no-auth) demand
  - Bandwidth limitations in serving data, connection restrictions for https access
  - Very limited CERN-based CPU; UNL, CERN VRE, EXPLORE at GoeGrid, … overview
  - Discussion around moving data to (≥1) US site / AmSC / FDP / HPDF / …
  - Also discussion of hosting data / metadata / links on HuggingFace
- LHCb's Ntuple Service (paper) is IMO a super cool idea; similar to OpenCosmo
- Monitoring via Motomo (website) and Monit-Grafana (open data portal, metadata)

Website visits in 2026



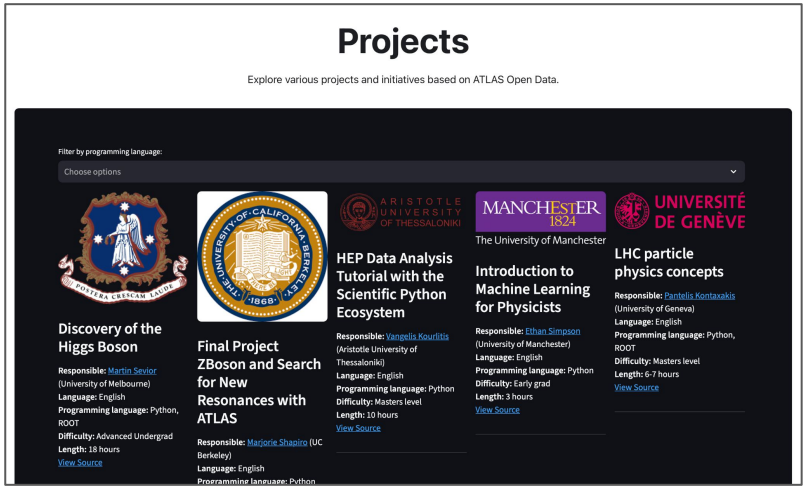
Portal access last 90 days

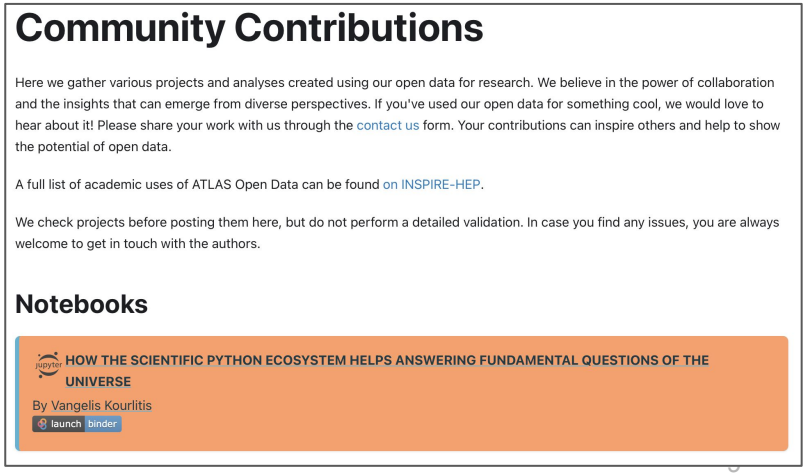
Additional Data Offerings

- In the last several weeks, multiple questions about additional data
  - RAW, delayed stream, TLA, PEB, parked, RAW, ‘complex’/‘detailed’, HL-LHC…
  - Generally speaking we make 10% of what we have available (resource limitations)
- Lots of discussion around data formats as well
  - Old belief was “simpler is better” and “avoid experiment software dependencies”
  - Now significant demand for complex data that would require experiment software
  - Bespoke datasets often in HDF5 (no love for the ROOT data loader, apparently)
- These are viable, but they take work and need prioritization (or **lots** more effort)

Getting help / community support

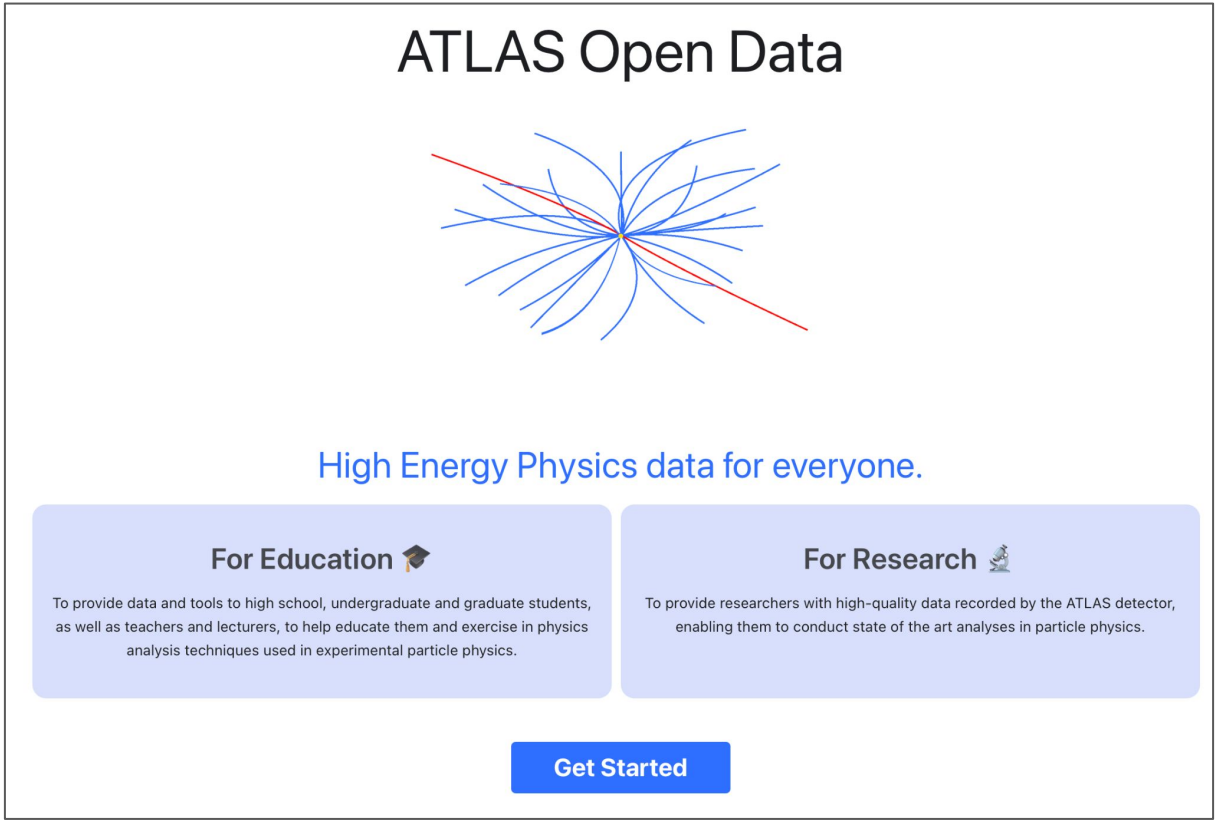
- We're very **happy to help**
- The CERN Open Data Forum is the **best place to ask questions**
  - Other users can find questions similar to theirs
  - We can easily crowdsource answers
  - We can keep track of how many people are asking great questions about the Open Data (we are asked about this regularly)
  - We have (obviously) discussed ChatBot support as well, and would be happy to have help there
- Let us know about your projects as well!
  - We are happy to hear about educational / outreach projects and research projects
  - You can also use those as inspiration… and we can **advertise** for you!





Summary / Conclusions

- We have lots of Open Data available for use
  - Both ‘big datasets’ and smaller custom datasets; offerings limited by effort
  - Read: we have not encountered any resistance inside the collaboration yet to the principle of releasing more data when the requests are *reasonably well justified*
- Monitoring is showing us what’s being used; surveys showing us how it lands
  - Fresh white paper on Open LHC Monte Carlo Event Generation, including AI/ML uses/cases
- Gathering contributions; happy to advertise your projects
- Lots of interesting opportunities for AI/ML use “for the open data” and “for the field”
- Upcoming white paper: “A vision for open data in the age of AI”
  - Intended to capture many of these opportunities



Thank you!

BERKELEY LAB

Big Open Data Datasets

For Research

Released

Proton-Proton Collisions

36 fb^{-1} proton-proton collisions in PHYSLITE format. 7B data, 2B MC evts, 373 MC sets. ~65 TB [link]

Released

Heavy Ion Minimum Bias

486 μb^{-1} lead-lead collisions in HION14 format. 221M data + 100k MC events. ~4 TB [link]

Released

Event Generation Data

>12B events, >6000 datasets in HEPMC format. 13 TeV and 13.6 TeV. ~1 PB [link]

Coming soon

Heavy Ion Hard Probes

New format, similar to PHYSLITE, being prepared now

For Education and Outreach

Released

8 TeV (2016)

3 fb^{-1} proton-proton collisions in XML and ROOT NTuple formats. 60M events, 42 MC sets [link]

Released

13 TeV (2020)

10 fb-1 proton-proton collisions in ROOT NTuple format. 940M events, 228 datasets [link]

Released

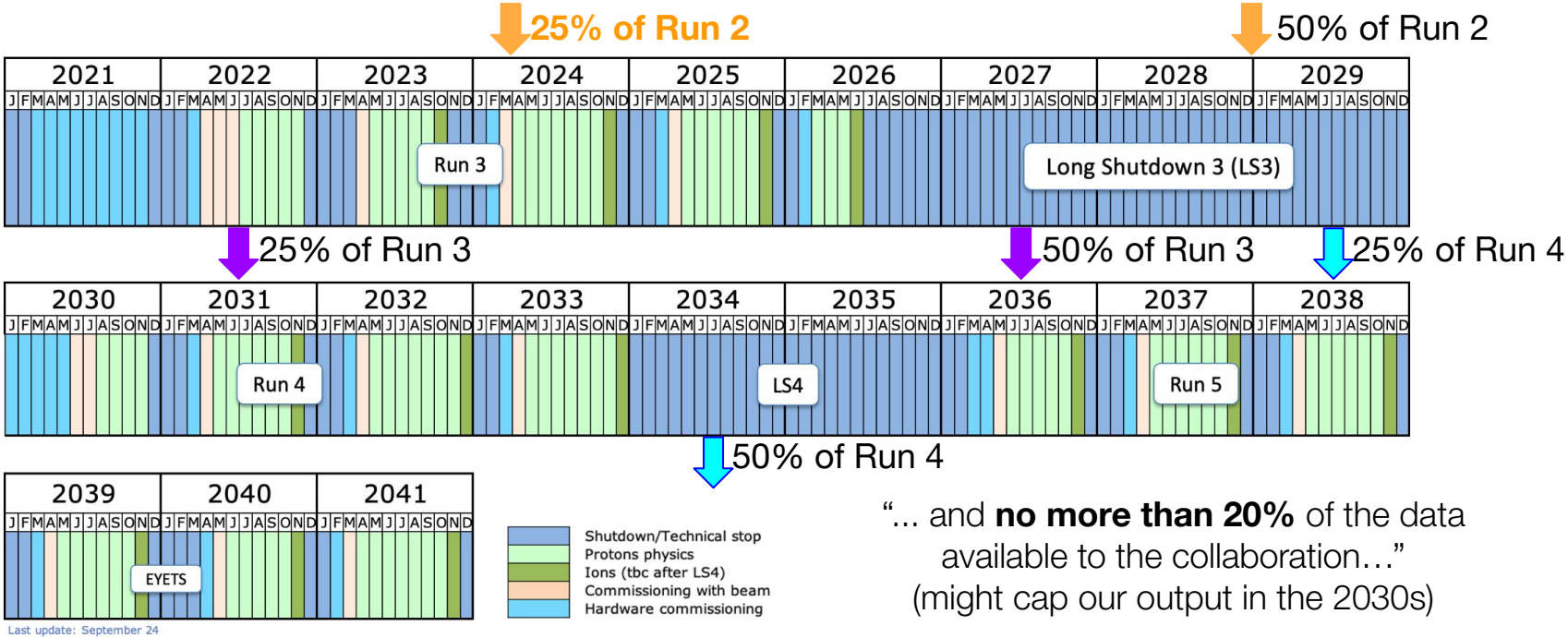
13 TeV (2025 Beta)

36 fb^{-1} proton-proton collisions in ROOT NTuple format. 9.8B events (dupes), 373 MC sets. ~2.5 TB [link]

All data are released under a CC0 license (free use) with a request for citation.
ATLAS members **may publish** using the Open Data.

Open Data for Research Schedule

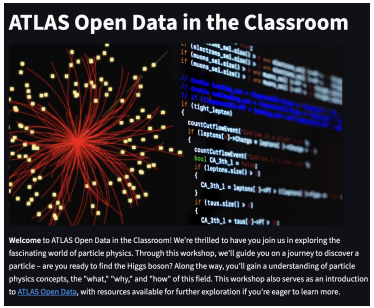
- ATLAS Schedule for Open Data releases



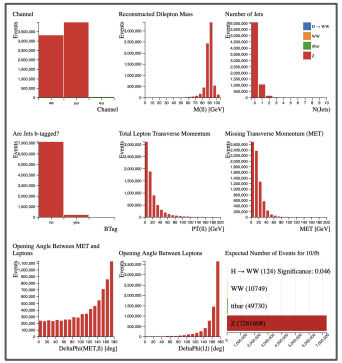
"... and **no more than 20%** of the data available to the collaboration…"
(might cap our output in the 2030s)

Our Documentation Goal

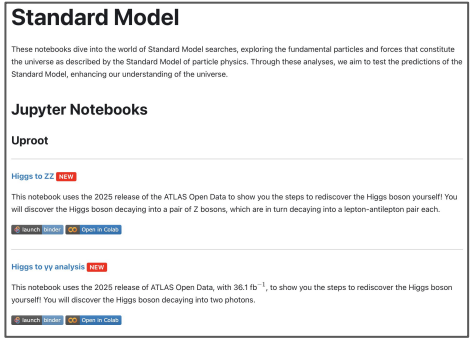
- Our goal is to have open data that's accessible to a wide variety of audiences
- We want learners to be able to move to more or less complex setups painlessly
- For more complexity than the Outreach and Education Open Data allow, try the Open Data for Research! If O&E OD are too complex, try the Histogram Analyzer!



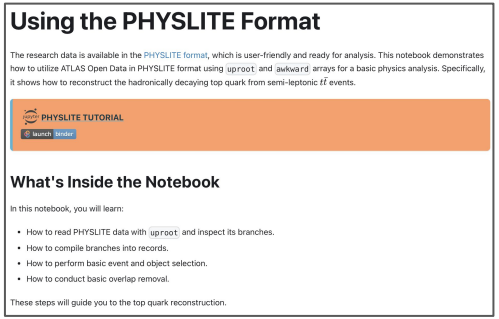
Introductory "Classroom Apps"
(Also in Spanish and Italian)



No-code "Histogram Analyzers"



Open Data for
Education and Outreach



Open Data for
Research

ML 'explainer'

What're the Metadata Like?

- Metadata are super important. Notebook tutorial. Runs on Colab, Binder, SWAN.
- Welcome to atlasopenmagic 🪄

```
[ ]: # First we install atlasopenmagic into our environment
     %pip install atlasopenmagic
```

```
[2]: # Now we can safely import atlasopenmagic
     import atlasopenmagic as atom
```

```
[3]: # Now let's see what releases are available to us
     atom.available_releases()
```

ATLAS Open Magic 🪄📊

Tests passing | pypi 1.8.0 | codecov 100%

`atlasopenmagic` is a Python package made to simplify working with ATLAS Open Data by providing utilities to manage metadata and URLs for streaming the data.

```
Available releases:
=======================================
2016e-8tev          2016 Open Data for education release of 8 TeV proton-proton collisions (https://opendata.cern.ch/record/3860).
2020e-13tev         2020 Open Data for education release of 13 TeV proton-proton collisions (https://cern.ch/2r7xt).
2024r-pp            2024 Open Data for research release for proton-proton collisions (https://opendata.cern.record/80020).
2024r-hi            2024 Open Data for research release for heavy-ion collisions (https://opendata.cern.ch/record/80035).
2025e-13tev-beta    2025 Open Data for education and outreach beta release for 13 TeV proton-proton collisions (https://opendata.cern.ch/record/93910).
2025r-evgen-13tev   2025 Open Data for research release for event generation at 13 TeV (https://opendata.cern.ch/record/160000).
2025r-evgen-13p6tev 2025 Open Data for research release for event generation at 13.6 TeV (https://opendata.cern.ch/record/160000).
```

```
[12]: # And let's use the latest release of Event Generation Open Data
      atom.set_release('2025r-evgen-13p6tev')
```

```
Fetching and caching all metadata for release: 2025r-evgen-13p6tev...
Fetched 1509 datasets so far...
Successfully cached 1509 datasets.
Active release: 2025r-evgen-13p6tev. (Datasets path: REMOTE)
```

What're the Metadata Like?

- Metadata are super important. Notebook tutorial. Runs on Colab, Binder, SWAN.
- Welcome to atlasopenmagic 🪄

```
[14]: # Now we can look at the metadata for a specific sample
      atom.get_metadata(510203)
      # Notice that the function here will accept either the dataset identifier or the
      # "physics short", a short unique descriptor for the sample %% [markdown] That's
      # a lot of metadata! Let's go through the fields a bit:
```

```
[14]: {'dataset_number': '510203',
 'physics_short': 'MGPy8EG_A14NNPDF30_SM4topsLOInclusive_run3',
 'e_tag': None,
 'cross_section_pb': 0.0092591,
 'genFiltEff': 1.0,
 'kFactor': 1.0,
 'nEvents': 10000,
 'sumOfWeights': None,
 'sumOfWeightsSquared': None,
 'process': None,
 'generator': 'MadGraph(v.3.5.3.atlas4)+Pythia8(v.310)+EvtGen(v.2.2.1)',
 'keywords': ['4top', 'Systematic', 'lo', 'sm', 'top'],
 'description': 'Standard-Model 4tops production at LO with MadGraph5 and Pythia8',
 'job_path': 'https://gitlab.cern.ch/atlas-physics/pmg/mcjoboptions/-/blob/master/510xxx/510203/mc.MGPy8EG_A14NNPDF30_SM4topsLOInclusive_run3.py',
 'CoMEnergy': 13600.0,
 'GenEvents': 41240000,
 'GenTune': 'A14 NNPDF23LO',
 'PDF': 'NULL',
 'Release': 'AthGeneration_23.6.24',
 'Filters': '',
 'cross_section_uncertainty': 0.0,
 'hepmc_version': 2,
 'release': {'name': '2025r-evgen-13p6tev'}}
```

That's a lot of metadata! Let's go through the fields a bit:

- `dataset_number` : Unique identifier assigned to each dataset.
- `physics_short` : Short name with information regarding the content of the dataset.
- `cross_section_pb` : Represents the probability of a particular interaction occurring, measured in picobarns (pb). It is a fundamental parameter that helps understanding the likelihood of specific particle interactions under given conditions.
- `genFiltEff` : Measure of the effectiveness of the selection criteria applied to the data. It indicates the fraction of events that pass the filters applied during the data processing stages.
- `kFactor` : Multiplicative correction factor used to account for higher-order effects in theoretical calculations. It adjusts the leading-order theoretical predictions to better match the observed data by incorporating next-to-leading order (NLO) or next-to-next-to-leading order (NNLO) corrections.

ATLAS Open Magic 🪄📊

Tests passing | pypi 1.8.0 | codecov 100%

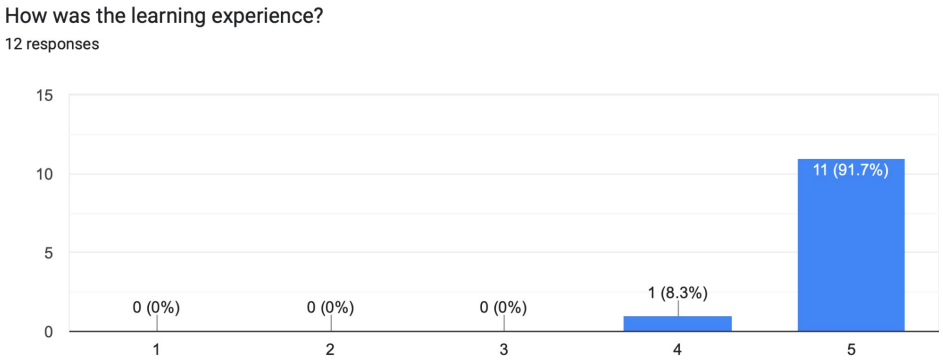
`atlasopenmagic` is a Python package made to simplify working with ATLAS Open Data by providing utilities to manage metadata and URLs for streaming the data.

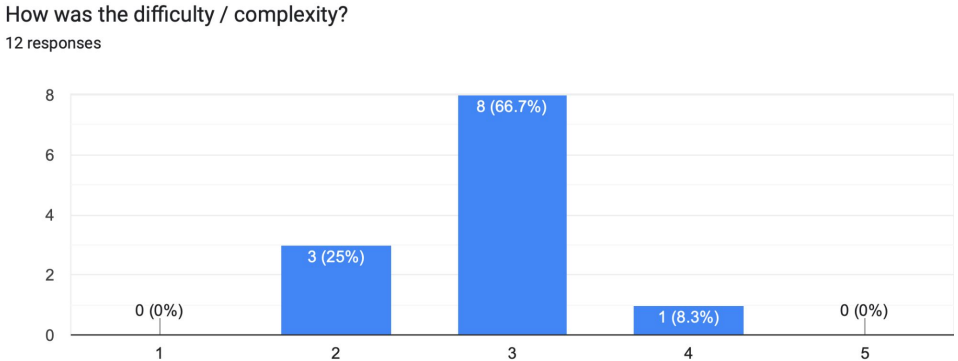
Considering tying this more closely together with our existing metadata systems (e.g. AMI and Rucio)

Now available: atlasopenmagic-mcp
NB: ami-mcp and rucio-mcp exist

Monitoring and Feedback

- Monitoring via Motomo (website) and Monit-Grafana (open data portal, metadata)
  - Still iterating with CERN IT on the best (long-term) metrics consistent with privacy rules
- Added some feedback forms throughout our materials so that we can see what folks think is working / isn't working, who we're reaching, etc
- We've implemented feedback forms in our notebooks and web applications
  - Very positive feedback, with most material at the right level (or a little easy)
- Most material has been pretty well received
  - Upcoming white paper from the LHC REI WG on event generation open data, highlighting use of our samples





AI/ML might be the future?

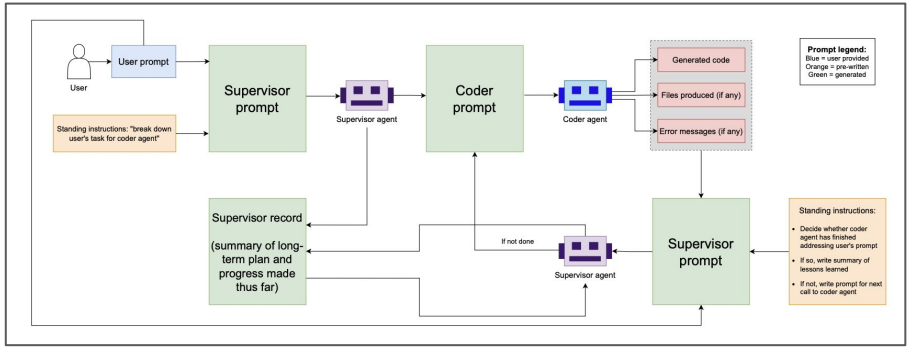
- The open data for education provides a great testing ground for many things
- Particularly for agentic workflow development and proof-of-concept work, no need to go straight to the research OD
- Interesting possibilities for Open Data and Analysis Preservation
  - Take this paper, create a workbook analysis for a second-year undergraduate, write the instructions in Hindi.
  - Produce a preserved workflow from this repo. Validate it from the INT note / L1 open data. Update it for the new release.
  - Could be game-changing for the effort problems we have in these areas.

Automating High Energy Physics Data Analysis with LLM-Powered Agents

Eli Gendreau-Distler,[1,2,*] Joshua Ho,[1,2,†] Dongwon Kim,[1,2,‡] Luc Tomas Le Pottier,[1,2,§] Haichen Wang,[1,2,¶] and Chengxi Yang[1,2,**]

[1] *Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA*
[2] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

We present a proof-of-principle study demonstrating the use of large language model (LLM) agents to automate a representative high energy physics (HEP) analysis. Using the Higgs boson diphoton cross-section measurement as a case study with ATLAS Open Data, we design a hybrid system that combines an LLM-based supervisor–coder agent with the `Snakemake` workflow manager.



VI. LIMITATIONS

This study shows that LLMs can support HEP data analysis workflows by interpreting natural language, generating executable code, and applying basic self-correction.

What Sorts of Things are There?

Concepts

These notebooks introduce a variety of concepts in High Energy Particle physics. They are intended to provide both a conceptual introduction and some real examples of how to study the concepts with the ATLAS Open Data.

Jupyter Notebooks

All New!

Uproot

Accessing Metadata NEW

This notebook introduces the atlasopenmagic package, which is used throughout these notebooks for environment setup and data access. It introduces the concept of metadata, explains what metadata are available for the samples that have been provided, and introduces the use of a handful of convenient search functions to identify samples that might be of interest based on their metadata.

launch binder Open in Colab

Detector Acceptance and Efficiency NEW

This notebook introduces the concepts of detector acceptance and efficiencies, used to understand the response of the detector to different types of events. Using the 2025 release of ATLAS Open Data, we guide you through a practical example of how to calculate acceptance and efficiency, and show you how these concepts are used in real physics analyses.

launch binder Open in Colab

Systematic Uncertainties NEW

This notebook introduces the concept of systematic uncertainties, providing a small variety of examples of uncertainties that you might run into while using the ATLAS Open Data. There is both a pedantic introduction and a demonstration of the calculation of uncertainties using the Open Data.

launch binder Open in Colab

Working with the Open Data Work in Progress

This notebook introduces the basics of how to work with the Open Data for Education and Outreach.

launch binder Open in Colab

Non-Collision Backgrounds NEW

This notebook introduces non-collision backgrounds — things that ATLAS records that aren't proton-proton (or heavy ion) collisions.

launch binder Open in Colab

Fluctuations NEW

This notebook is designed to help build some intuition around fluctuations: what level of agreement we expect between our background estimates and the data even in an analysis where we have done everything right, because of the intrinsic limitations of statistical and systematic uncertainties.

launch binder Open in Colab

Standard Model

These notebooks dive into the world of Standard Model searches, exploring the fundamental particles and forces that constitute the universe as described by the Standard Model of particle physics. Through these analyses, we aim to test the predictions of the Standard Model, enhancing our understanding of the universe.

Jupyter Notebooks

Revised!

Uproot

Higgs to ZZ NEW

This notebook uses the 2025 release of the ATLAS Open Data to show you the steps to rediscover the Higgs boson yourself! You will discover the Higgs boson decaying into a pair of Z bosons, which are in turn decaying into a lepton-antilepton pair each.

launch binder Open in Colab

Higgs to γγ analysis NEW

This notebook uses the 2025 release of ATLAS Open Data, with 36.1 fb^{-1}, to show you the steps to rediscover the Higgs boson yourself! You will discover the Higgs boson decaying into two photons.

launch binder Open in Colab

Find the Z boson NEW

This notebook guides you through finding the Z boson in events with two muons. It gives you a variety of possible extensions to explore these events, understand more about the Z boson, identify other Standard Model particles, or search for new particles.

launch binder Open in Colab

Higgs boson to Muon-Antimuon Pair NEW

Using the 2025 release of ATLAS Open Data, this notebook walks you through the process of rediscovering Higgs-boson production via its decay into a muon–antimuon pair..

launch binder Open in Colab

Searching for top-antitop quark pairs NEW

This notebook uses 2025 release of the ATLAS Open Data to guide you through the steps needed to rediscover the production of top-antitop quark pairs.

launch binder Open in Colab

New!

WZ to three leptons NEW

This notebook uses the 2025 release of ATLAS Open Data, with 36.1 fb^{-1}, to show you the steps to find events where a W- and a Z-boson have been produced and decayed into lepton-neutrino and lepton-antilepton pairs, respectively! You will be able to reconstruct the mass of both the W- and the Z-boson using both electrons and muons.

launch binder Open in Colab

Higgs to bb analysis NEW

This notebook uses the 2025 release of ATLAS Open Data, with 36.1 fb^{-1}, to show you the steps to attempt to find the Higgs boson when it decays into two b-quarks. This is a challenging analysis even for ATLAS

Statistical and Systematic Fluctuations

This notebook uses ATLAS Open Data 2025 beta release to build your intuition for statistical and systematic fluctuations. It is intended for an education audience and is written to be accessible to a wide range of students.

What are Fluctuations?

When we perform an analysis with ATLAS data, very often we are confronted with histograms of data and a background prediction, and we need to make a judgement about whether the two agree. It can be difficult to build intuition for what level of disagreement one might expect, and often even experienced physicists disagree about whether there might be a discrepancy that is worthy of closer inspection (for example, one that might be a hint towards a new discovery). The alternative, of course, is that the differences are simply statistical fluctuations — the regular thing we expect in every analysis.

Looking at fluctuations

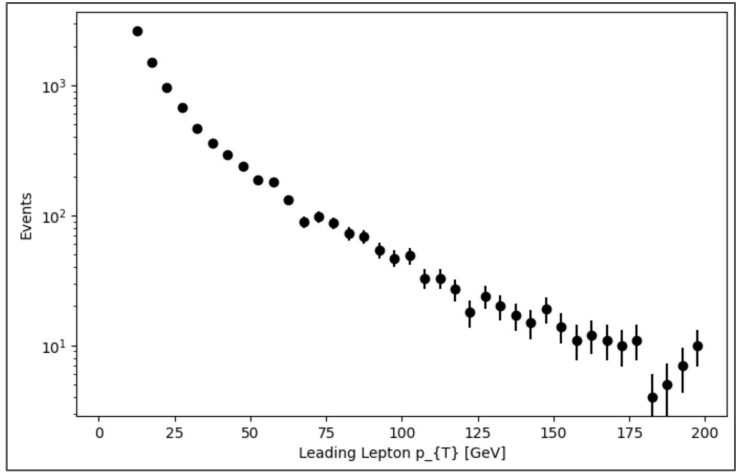
In this notebook, we are going to start from the background in the $H \to \gamma\gamma$ analysis, which you can find at the end of our $H \to \gamma\gamma$ notebook.

We are going to use a random number generator to create two kinds of fluctuations:

- Statistical fluctuations. These are random fluctuations that occur in all counting experiments. As a simple example, imagine counting the number of cars that pass by you on a road. You might have a very good understanding of the number of cars that will go by in an hour (the *rate*), but if you watch the road for only a few seconds it is likely that you would not see an "average" number of cars go by — it might be quite a bit lower or higher than you expect.
- Systematic fluctuations. Most data analyses have *systematic uncertainties*, which we explore in this notebook. These uncertainties cause a systematic bias when estimating the expected number of events we will see. We might always be a bit high, or a bit low.

Event Generation Open Data

- New, 12.7B events in ~6500 datasets
  - Access via atlasopenmagic (metadata tutorial)
  - Publicly documented naming, sample availability, metadata, limitations, how to combine… all useful for ATLAS newcomers as well!
- We provide a fully worked-out example notebook
  - Runs on Binder; some features not as nice on Colab
  - Set up, sample identification, files access
  - Visualizing events with pyhepmc
  - Making basic plots
  - Running Delphes, examining output
  - Takes <5 minutes to run
- Significant community interest in the Open Data
  - Upcoming white paper from the LHC REI WG on evgen open data, highlighting use of our samples




```
[16]: import uproot # for reading .root files
      # Get the tree with our data directly from the ROOT file
      tree = uproot.open("delphes_output.root:Delphes")
      # Just for an example here, we'll print the transverse momenta of the electrons
      tree["Electron.PT"].arrays()

[16]: [{'Electron.PT': []},
       {'Electron.PT': []},
       {'Electron.PT': []},
       {'Electron.PT': []},
       {'Electron.PT': []},
       {'Electron.PT': [75.4]},
       {'Electron.PT': []},
       {'Electron.PT': [39.1]},
       {'Electron.PT': []},
       {'Electron.PT': []},
       ...,
       {'Electron.PT': []},
       {'Electron.PT': [51.1]},
       {'Electron.PT': []},
       {'Electron.PT': []},
       {'Electron.PT': []},
       {'Electron.PT': [78.5]},
```

Other Open and Preserved Data News

- New bespoke datasets since the last DPHEP Workshop
  - Simulation voxelized photon showers dataset
  - ttbar for ML-based flavour tagging dataset
  - Several more in the pipe — mostly supporting AI/ML applications
- Still working to provide Plot records and HepData
  - Major milestone: in 2025, **all** exotics analyses had HepData! (100% coverage)
- Unfortunately, saw a drop in Rivet coverage, driven by the impressive number of analyses published since the last DPHEP workshop…
  - Seems to be universally true around the ring…

Last DPHEP Workshop

Rivet analysis coverage

Rivet analyses exist for 1838/6446 papers = 29%. 261 priority analyses required.

Total number of Inspire papers scanned = 10889, at 2024-08-08

Breakdown by identified experiment (in development):

Key	ALICE	ATLAS	CMS	LHCb	F
Rivet wanted (total):	380	477	562	205	1
Rivet REALLY wanted:	54	62	98	15	0
Rivet provided:	44/424 = 10%	212/689 = 31%	135/697 = 19%	71/276 = 26%	1 3

Rivet analysis coverage

Rivet analyses exist for 1637/6404 papers = 26%. 329 priority analyses required.

Total number of Inspire papers scanned = 11088, at 2025-08-08

Breakdown by identified experiment (in development):

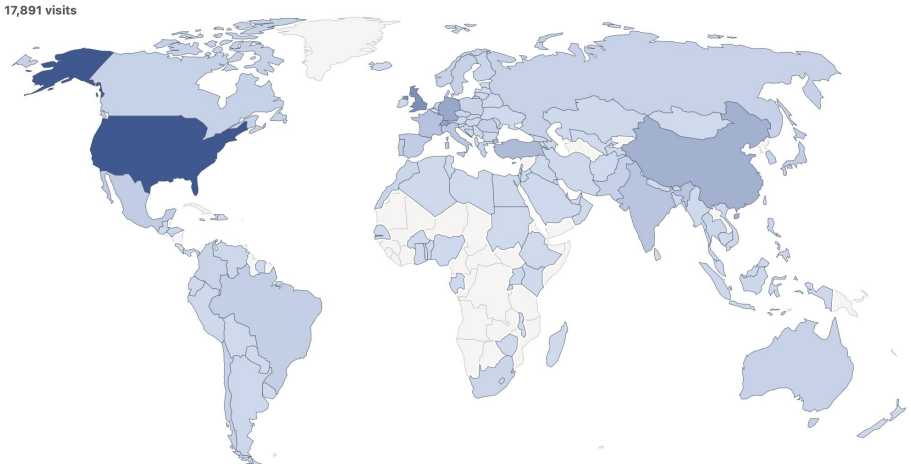
Key	ALICE	ATLAS	CMS	LHCb
Rivet wanted (total):	411	556	633	150
Rivet REALLY wanted:	87	78	106	19
Rivet provided:	40/451 = 9%	218/774 = 28%	136/769 = 18%	75/225 = 33%

Today

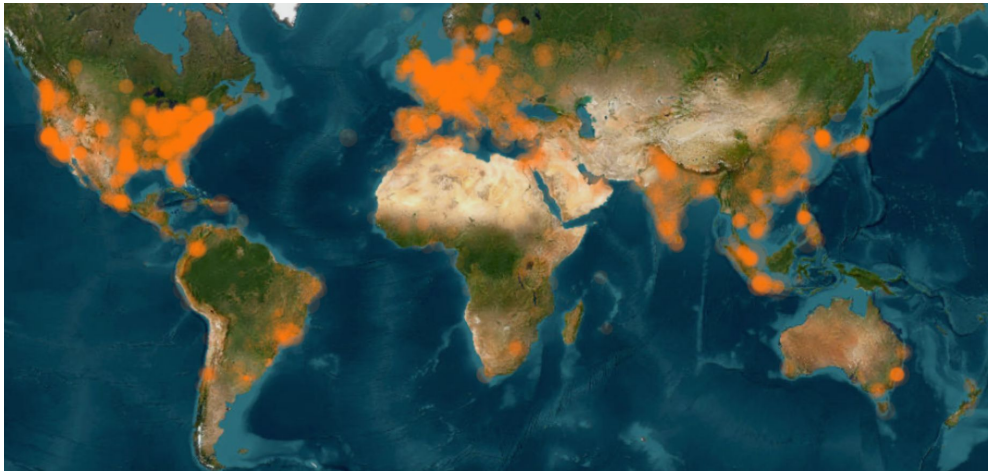
Newly-available Monitoring

- Monitoring via Motomo (website) and Monit-Grafana (open data portal, metadata)
  - Still iterating with CERN IT on the best (long-term) metrics consistent with privacy rules
- Added some feedback forms throughout our materials so that we can see what folks think is working / isn't working, who we're reaching, etc

Website visits in 2H 2025



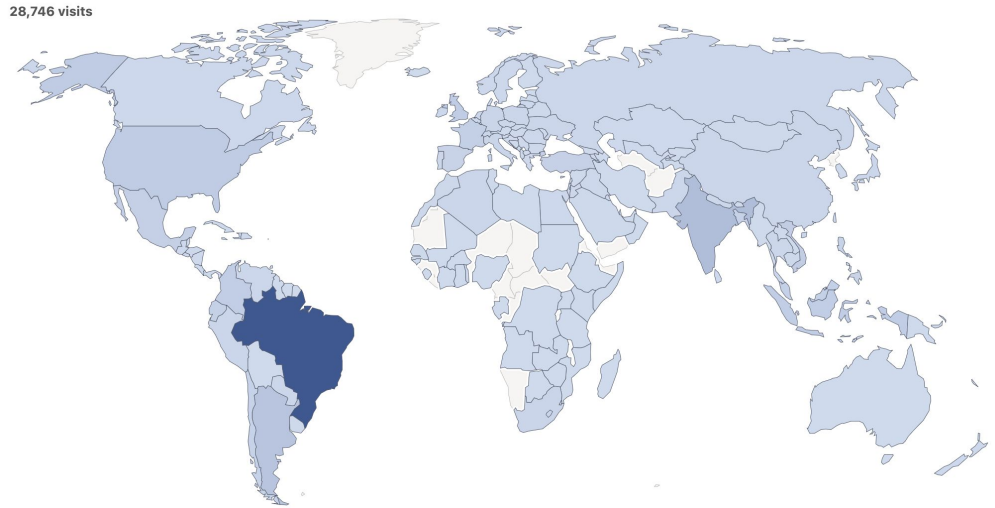
Portal access since September 2025



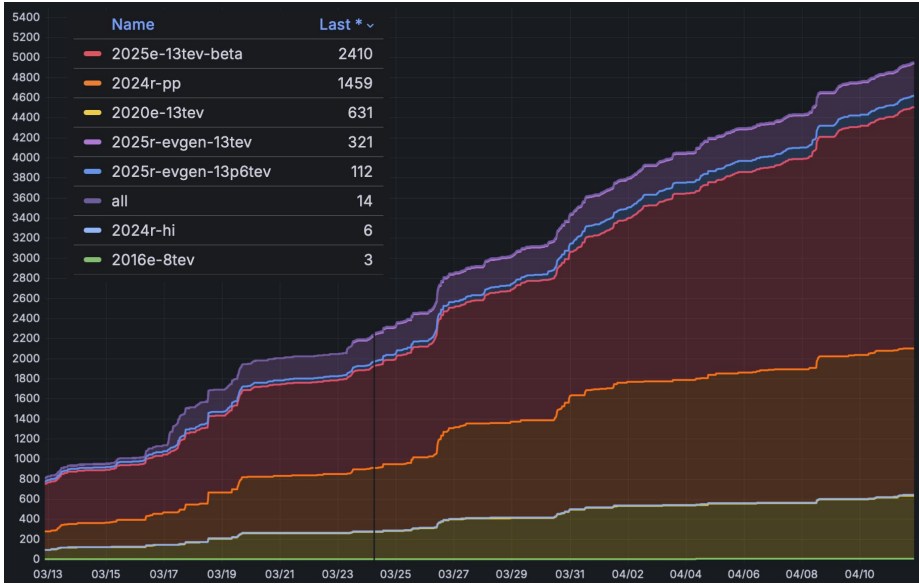
Newly-available Monitoring

- Monitoring via Motomo (website) and Monit-Grafana (open data portal, metadata)
  - Still iterating with CERN IT on the best (long-term) metrics consistent with privacy rules
- Added some feedback forms throughout our materials so that we can see what folks think is working / isn't working, who we're reaching, etc

Website visits in 2026



Metadata release setup, last 30 days



Updated Collaboration Policies

- Software Policy is now more strongly open source, recommending DOIs
  - *All* software internally visible; software open by default
- Data Preservation Policy now more firm about Run 1–3
  - We commit to **preserving** all raw data
  - We formally have *no active support* for Run 1 data
  - For Run 2 / 3 data during the HL-LHC era we expect to:
    - Allow statistical combinations with Run 4 / 5 (of course)
    - Retain ‘final’ analysis formats only (e.g. DAOD_PHYSLITE)
    - Produce new MC using fully containerized workflows
    - Drop support for ‘new’ processing of Run 2 / 3 data in our main software releases (reconstructing the data in a new release implies new calibrations and uncertainties, which require an enormous amount of effort)
- Documentation policy now public by default
  - Should mean less duplication of documentation for the Open Data users

(The Pantry in Los Angeles; the gold standard of “Open by default” for 90+ years — legend has it, they lost the key to the front door, it had been so long since they needed it)