CERN Open Data Portal

Pablo Saiz

IOP Open Data Workshop for Higher Education

University of Sussex, UK **14 May 2026**

CERN Open Data Portal

Quantitative Metrics
downloads, citations, …

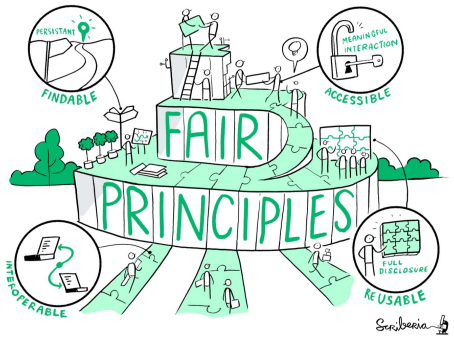
Qualitative Insights
case studies, stories, …

Equity & Inclusion
accessibility, diversity, …

How to measure impact

Open Data is good

FAIR Open Data is even better



eosc

European Open Science Cloud

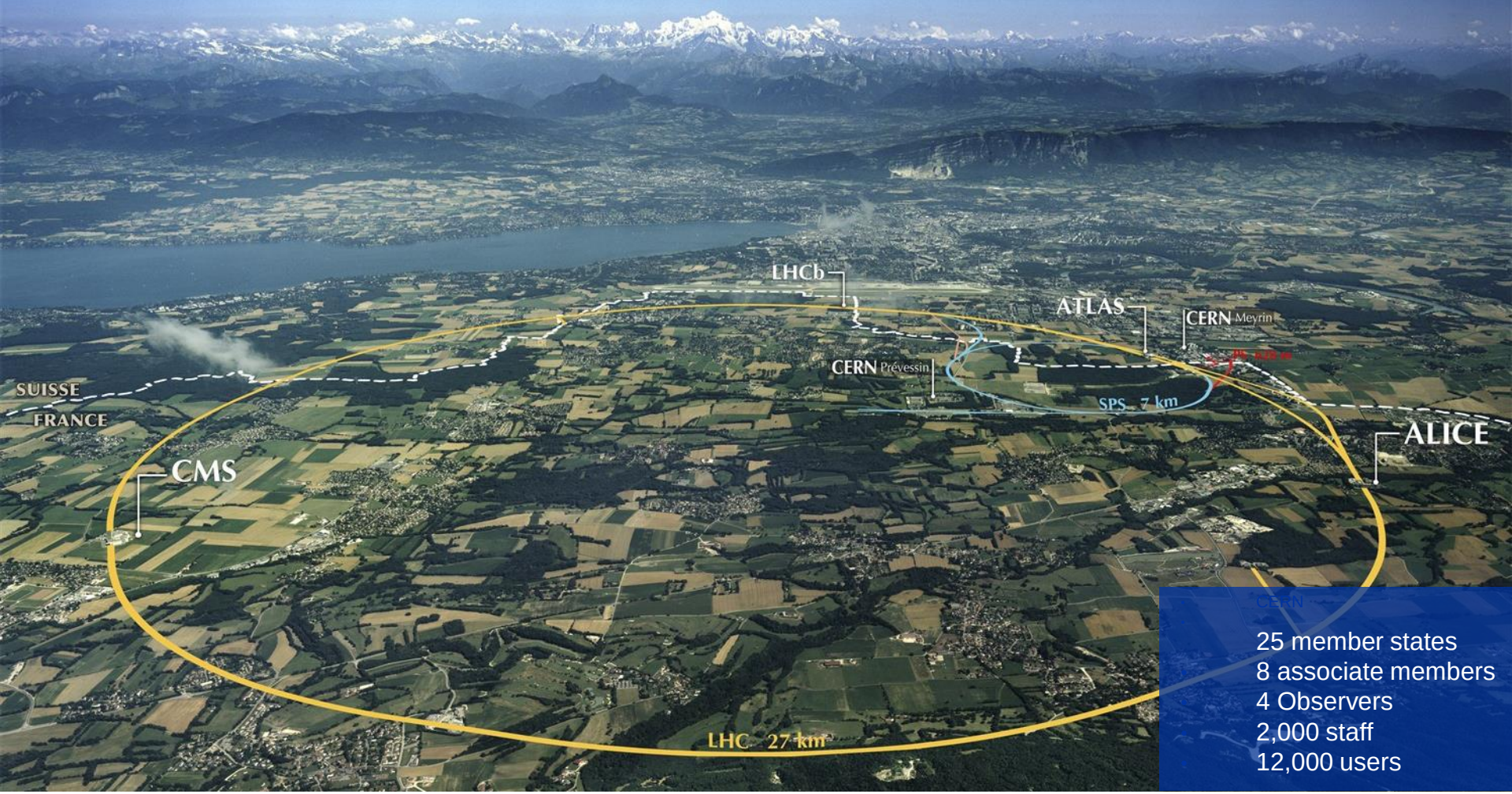
Other Fields

CERN

Organisation Européenne pour la Recherche Nucléaire

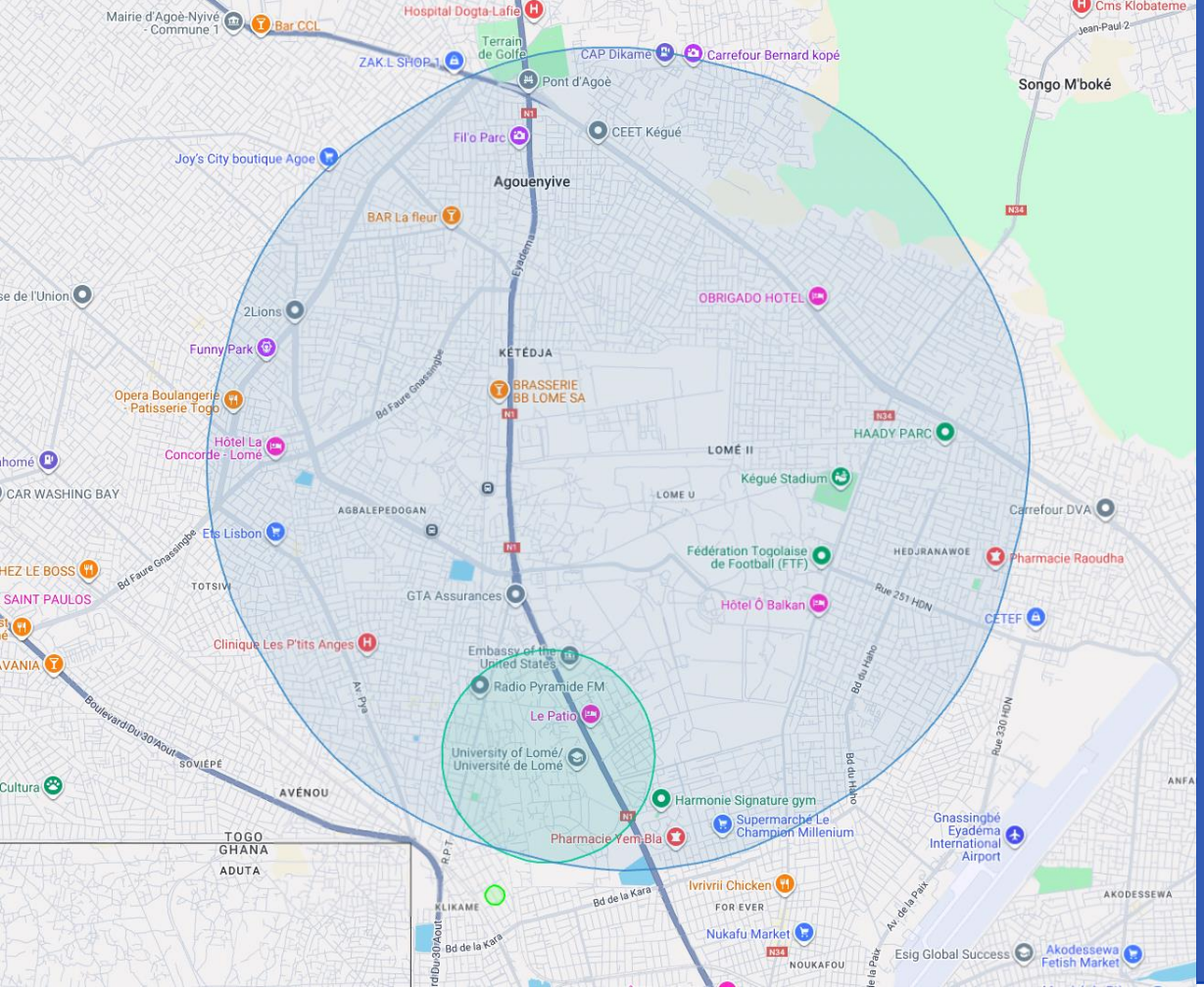
European Organization for Nuclear Research

1954



25 member states
8 associate members
4 Observers
2,000 staff
12,000 users

What if it was in Lomé?

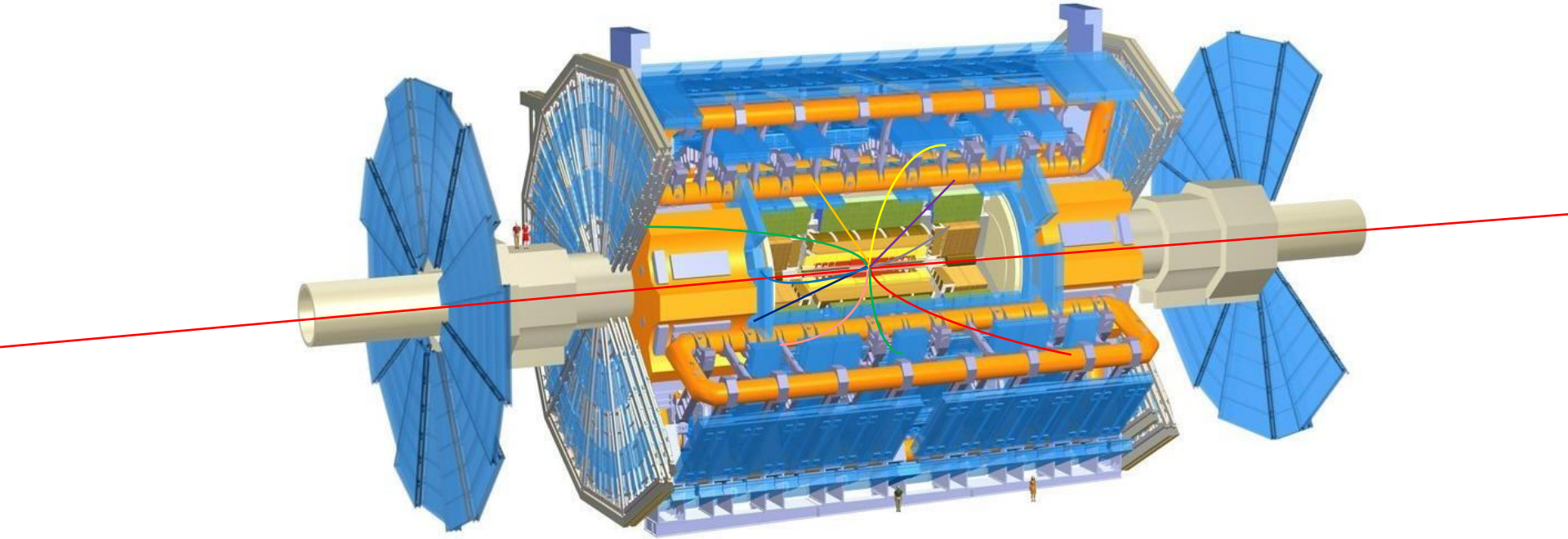


What if it was in Lomé?

And with the FCC

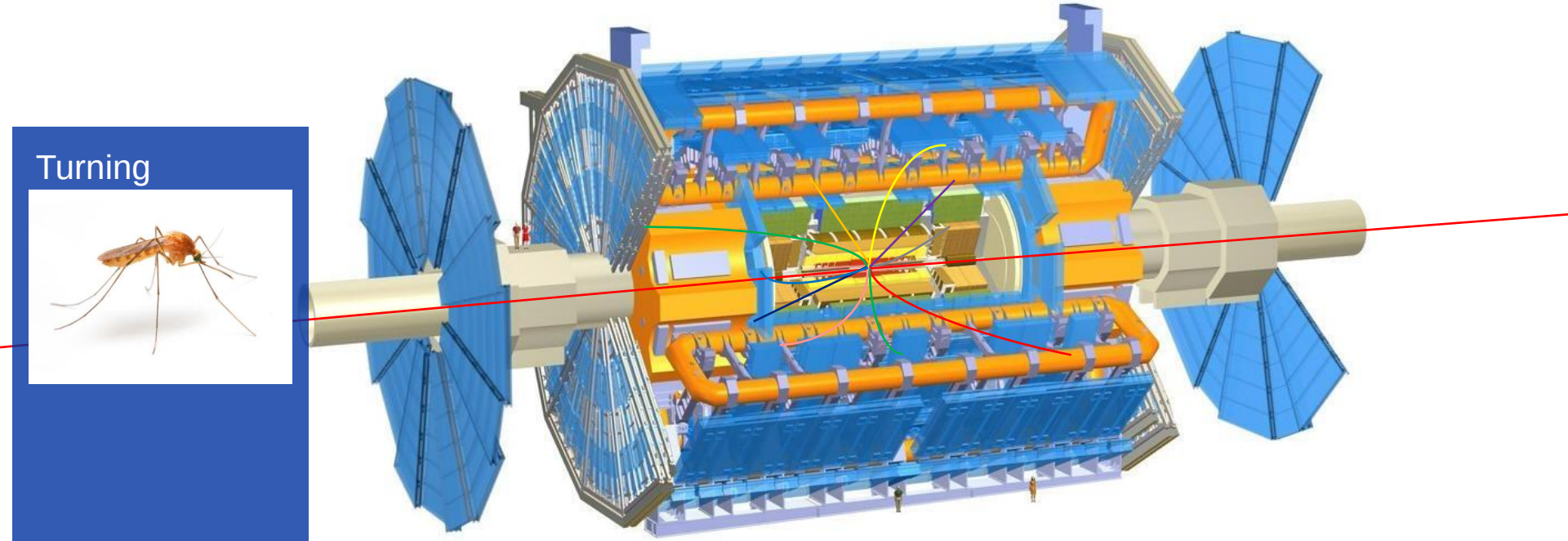
The most powerful magnets

The largest detectors



The largest detectors

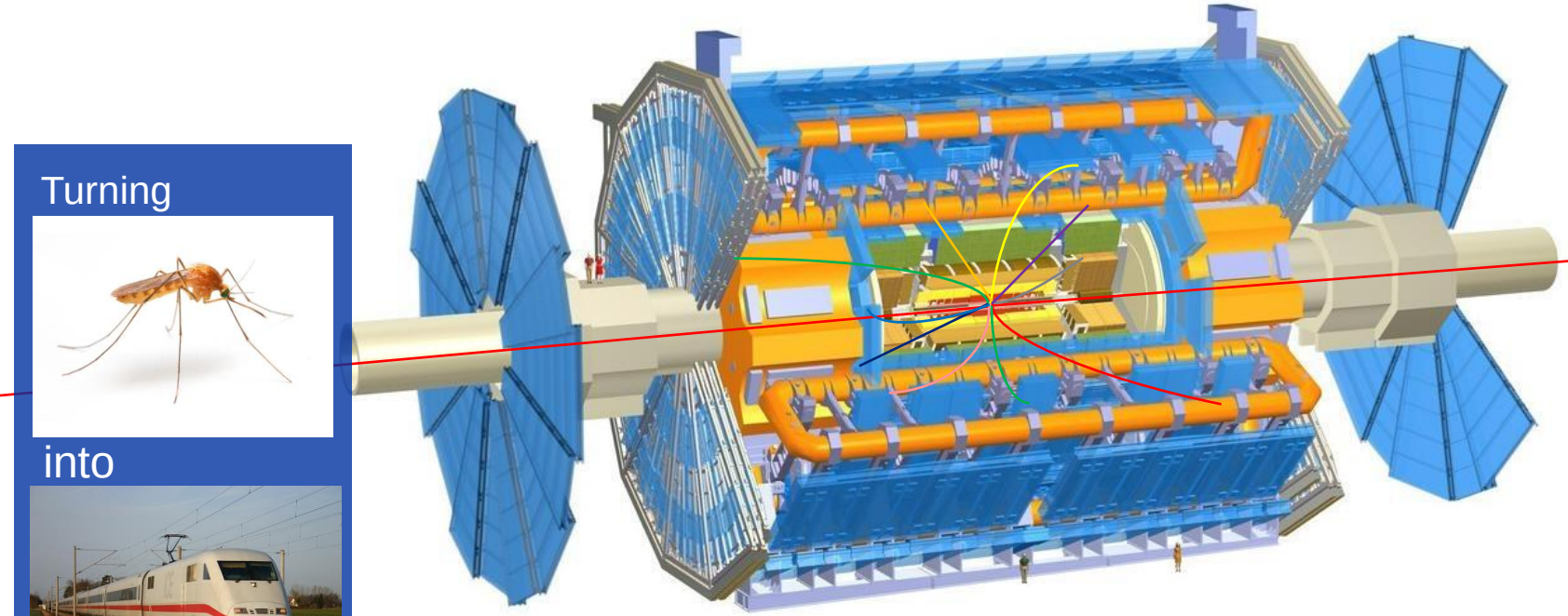
Turning



The largest detectors

Turning

into



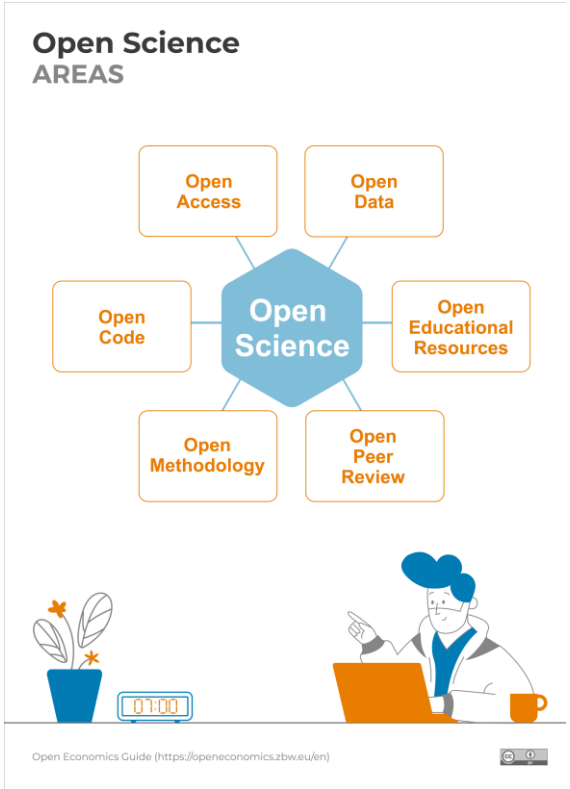
The biggest vaccum

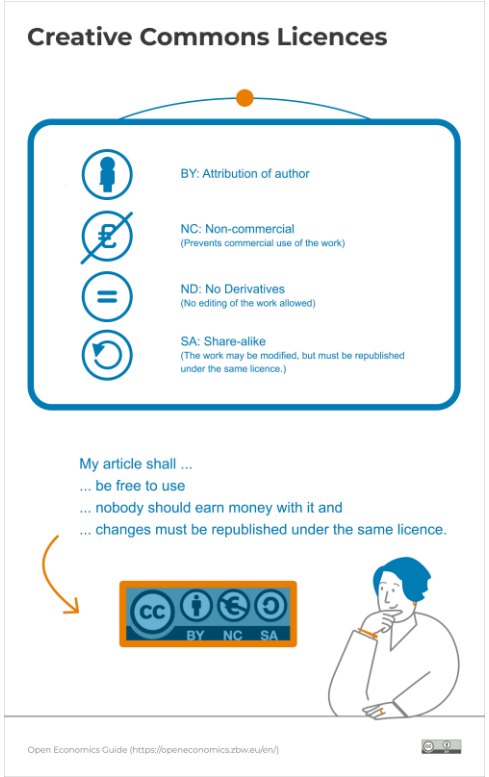

The coolest place

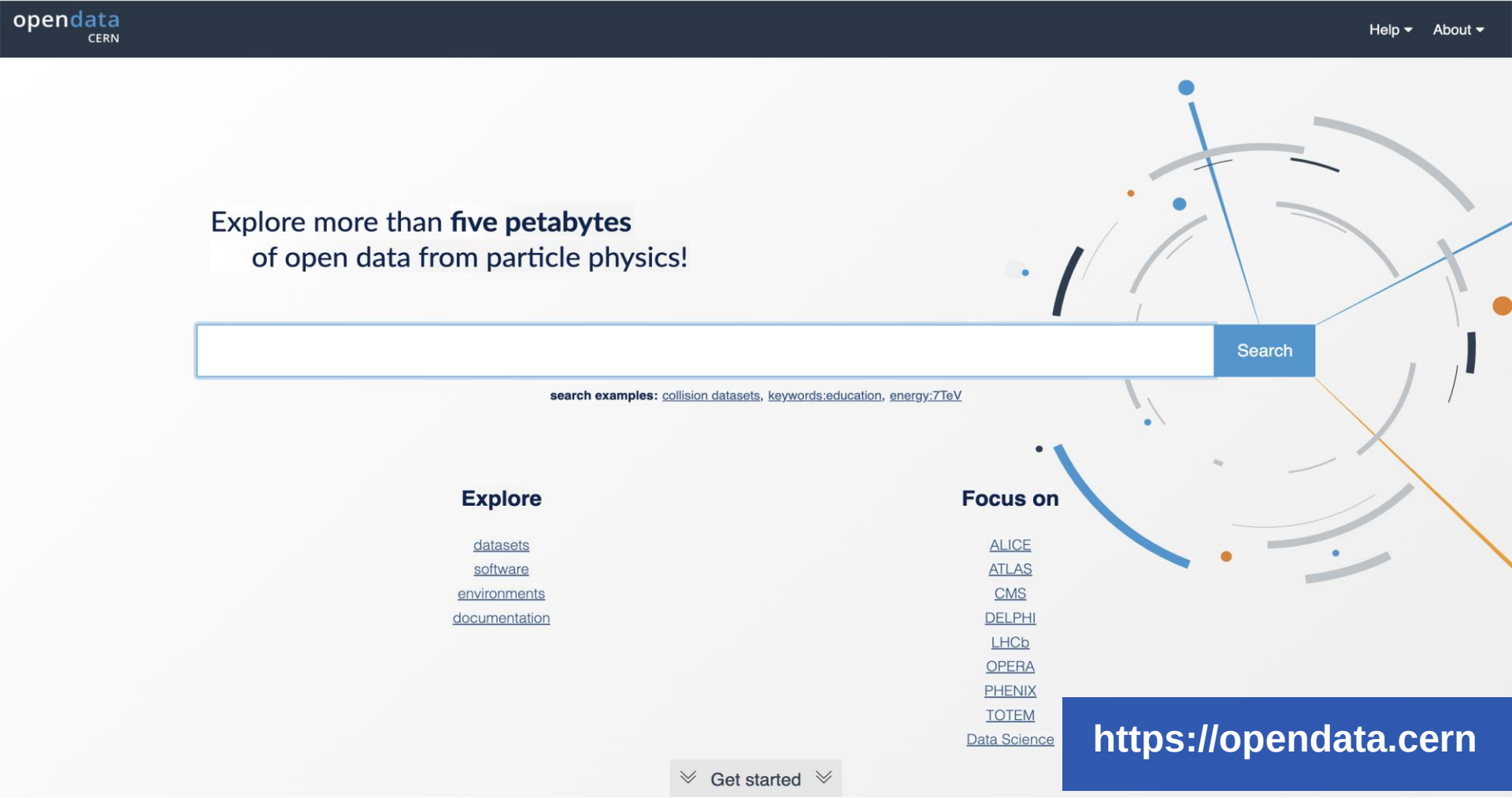
The biggest computing grid


SOCAR
CHANGE
CERN Open Science
zenodo
White Rabbit
SCOAP 3

Open Science






opendata
CERN
Help
About
Explore more than five petabytes
of open data from particle physics!
Search
search examples: collision datasets, keywords:education, energy:7TeV
Explore
datasets
software
environments
documentation
Focus on
ALICE
ATLAS
CMS
DELPHI
LHCb
OPERA
PHENIX
TOTEM
Data Science
Get started
https://opendata.cern

The Turing Way Community, & Scriberia. (2020)

Illustrations from the Turing Way book dashes. Zenodo. https://doi.org/10.5281/zenodo.3695300

opendata CERN

Help About

Type something Search

80,031 result(s) found

Sort by Most recent

Current parameters Clear all

Type

- Dataset (65,864)
  - Collision (756)
  - Derived (1,234)
  - Simulated (63,888)
- Documentation (7,251)
  - About (12)
  - Activities (20)
  - Authors (8)
  - Guide (40)
  - Help (3)
  - Policy (9)
  - Report (1)
  - Stripping (7,158)
- Environment (62)
  - Condition (20)
  - VM (18)
  - Validation (24)
- Glossary (1,006)
- News (29)
- Software (55)

ATLAS releases 2025 beta open data for education and outreach

The ATLAS Collaboration releases 25 datasets for education and outreach in a public 2025 "beta" release.

News ATLAS

ATLAS HEPMC format Specialised SM Z-boson 13 TeV

Specialised SM Z-boson samples at sqrt(s)=13 TeV from the ATLAS experiment. This record collects related samples. Some care is required when combining samples in order to ensure a complete physics ...

Dataset Simulated Standard Model ElectroWeak ATLAS

HEPMC format 13 TeV proton-proton Open Data from the ATLAS experiment

HEPMC format 13 TeV proton-proton Open Data from the ATLAS experiment.This summary record is provided as a single point of entry and single record to cite for the HEPMC open data.As additional data...

Dataset Simulated ATLAS

ATLAS $t\bar{t}$ simulation for ML-based jet flavour tagging (JetSet)

Flavour-tagging — the task of identifying heavy flavor jets — is essential for many physics analyses at the ATLAS experiment. This dataset, released for public use, can be used to train and evaluat...

Dataset Derived Simulated ATLAS

ATLAS ROOT ntuple format Run 2 2015+2016 proton-proton collision data beta release, 2J2LMET30 skim

Run 2 2015+2016 proton-proton collision data beta release, 2J2LMET30 skim from the ATLAS experiment

Dataset Collision ATLAS

VBF1Parked primary dataset in AOD format from Run of 2012 (/VBF1Parked/Run2012D-22Jan2013-v1/AOD)

/VBF1Parked/Run2012D-22Jan2013-v1/AOD, CMS collaboration

Cite as: CMS collaboration (2022). VBF1Parked primary dataset in AOD format from Run of 2012 (/VBF1Parked/Run2012D-22Jan2013-v1/AOD). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.S56K.7PUP

Dataset | Collision | CMS | 8TeV | pp | CERN-LHC

Description

VBF1Parked primary dataset in AOD format from RunD of 2012. Run period from run numbe

The list of validated runs, which must be applied to all analyses, can be found in

CMS list of validated runs Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSO

Dataset characteristics

334452423 events. **32618** files. **94.4 TiB** in total.

System details

Recommended global tag for analysis: FT53_V21A_AN6
Recommended release for analysis: CMSSW_5_3_32
Recommended container image for analyses is available in the following locations (see guid

- `docker.io/cmsopendata/cmssw_5_3_32-slc6_amd64_gcc472:latest`
- `gitlab-registry.cern.ch/cms-cloud/cmssw-docker-opendata/cmssw_5_`

How were these data selected?

Events in this primary dataset were selected because of the presence of at least two jets wit

Data taking / HLT
The collision data were assigned to different RAW datasets using the following HLT configura

Data processing / RECO
This primary AOD dataset was processed from the RAW dataset by the following step:
Step: RECO

How were these data validated?

During data taking all the runs recorded by CMS are certified as good for physics analysis if all subdetectors, trigger, lumi and physics objects (tracking, electron, muon, photon, jet and MET) show the expected performance. Certification is based first on the offline shifters evaluation and later on the feedback provided by detector and Physics Object Group experts. Based on the above information, which is stored in a specific database called Run Registry, the Data Quality Monitoring group verifies the consistency of the certification and prepares a json file of certified runs to be used for physics analysis. For each reprocessing of the raw data, the above mentioned steps are repeated. For more information see:

CMS data quality monitoring: Systems and experiences

The CMS Data Quality Monitoring software experience and future improvements

The CMS data quality monitoring software: experience and future prospects

How can you use these data?

You can access these data through the CMS Open Data container or the CMS Virtual Machine. See the instructions for setting up one of the two alternative environments and getting started in

Running CMS analysis code using Docker

How to install the CMS Virtual Machine

Getting started with CMS open data

Files and indexes

Index description	Index size	
VBF1Parked AOD dataset file index (1 of 70) for access to data	3.8 TiB	List files / Download index
VBF1Parked AOD dataset file index (2 of 70) for access to data	3.8 TiB	List files / Download index
VBF1Parked AOD dataset file index (3 of 70) for access to data	3.9 TiB	List files / Download index
VBF1Parked AOD dataset file index (4 of 70) for access to data	4.1 TiB	List files / Download index
VBF1Parked AOD dataset file index (5 of 70) for access to data	4.2 TiB	List files / Download index

« ‹ 1 2 3 4 5 ... 14 › »

HZZAnalysis.ipynb

File Edit View Insert Runtime Tools Help

Share

Commands + Code + Text Run all Copy to Drive Connect

How to rediscover the Higgs boson yourself!

open data ATLAS

This notebook uses ATLAS Open Data http://opendata.atlas.cern 2025 beta release to show you the steps t
yourself!

ATLAS Open Data provides open access to proton-proton collision data at the LHC for educational purpos
are ideal for high-school, undergraduate and postgraduate students.

Notebooks are web applications that allow you to create and share documents that can contain for exampl

1. live code
2. visualisations
3. narrative text

What is the Higgs boson?

The Higgs boson is a fundamental particle predicted by the Standard Model. It is a manifestation of the Hig
the fundamental particles. However, it is incredibly hard to produce. At the LHC, a Higgs particle is produce
collisions! This tiny fraction makes it very difficult to detect. Nevertheless, after years of data collection, the
discovered in 2012 by CMS and ATLAS experiments at CERN. In this tutorial, we shall be following their exa

Detecting the Higgs

This analysis loosely follows the paper on the discovery of the Higgs boson by ATLAS (mostly Section 4 and

The Higgs boson can be produced in many different ways. In particle physics, we describe these productio
diagrams. These diagrams allow us to visualise particle processes while also acting as powerful tools for ca
information on Feynman diagrams.

There are four main production modes of the Higgs boson, and the respective Feynman diagrams:

We will use the atlasopenmagic to access the open data directly from the ATLAS OpenData Portal so no need to download any samples.
First we need to install the package

Import the module and load the release.

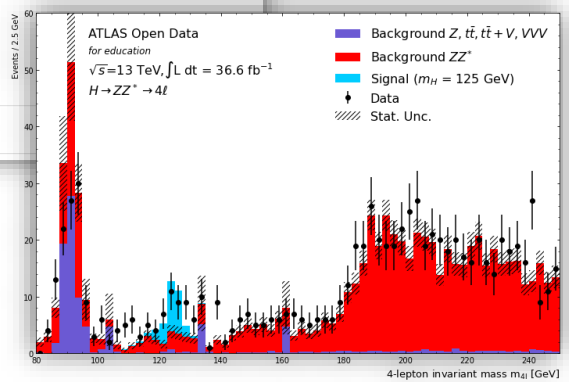
```
[ ]
import atlasopenmagic as atom
atom.available_releases()
atom.set_release('2025e-13tev-beta')
```

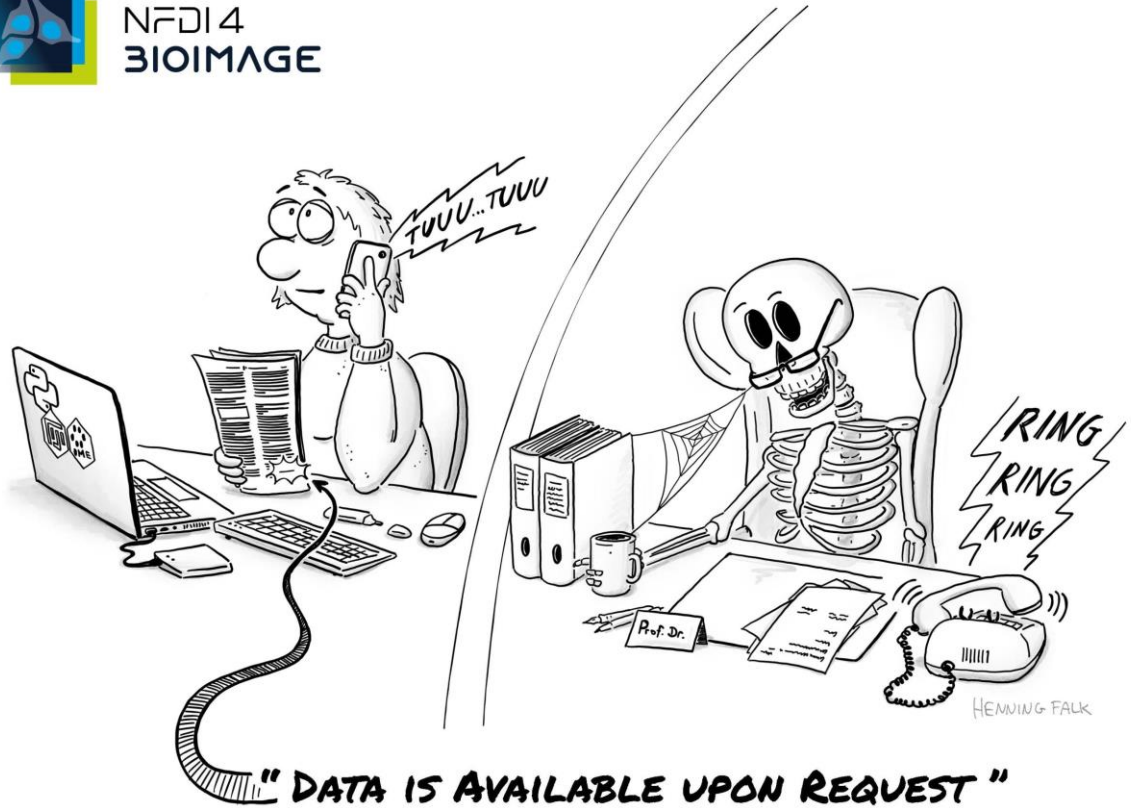
```
Available releases:
==========================================
2016e-8tev        2016 Open Data for education release of 8 TeV proton-proton collisions (https://opendata.cern.ch/record/38
2020e-13tev       2020 Open Data for education release of 13 TeV proton-proton collisions (https://cern.ch/2r7xt).
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://opendat
2025r-evgen       2025 Open Data for research release for event generation (https://opendata.cern.ch/record/160000).
Fetching and caching all metadata for release: 2025e-13tev-beta...
Successfully cached 374 datasets.
Active release: 2025e-13tev-beta. (Datasets path: REMOTE)
```

Example 1: Reading data

We would like to read some of the data from the open dataset.

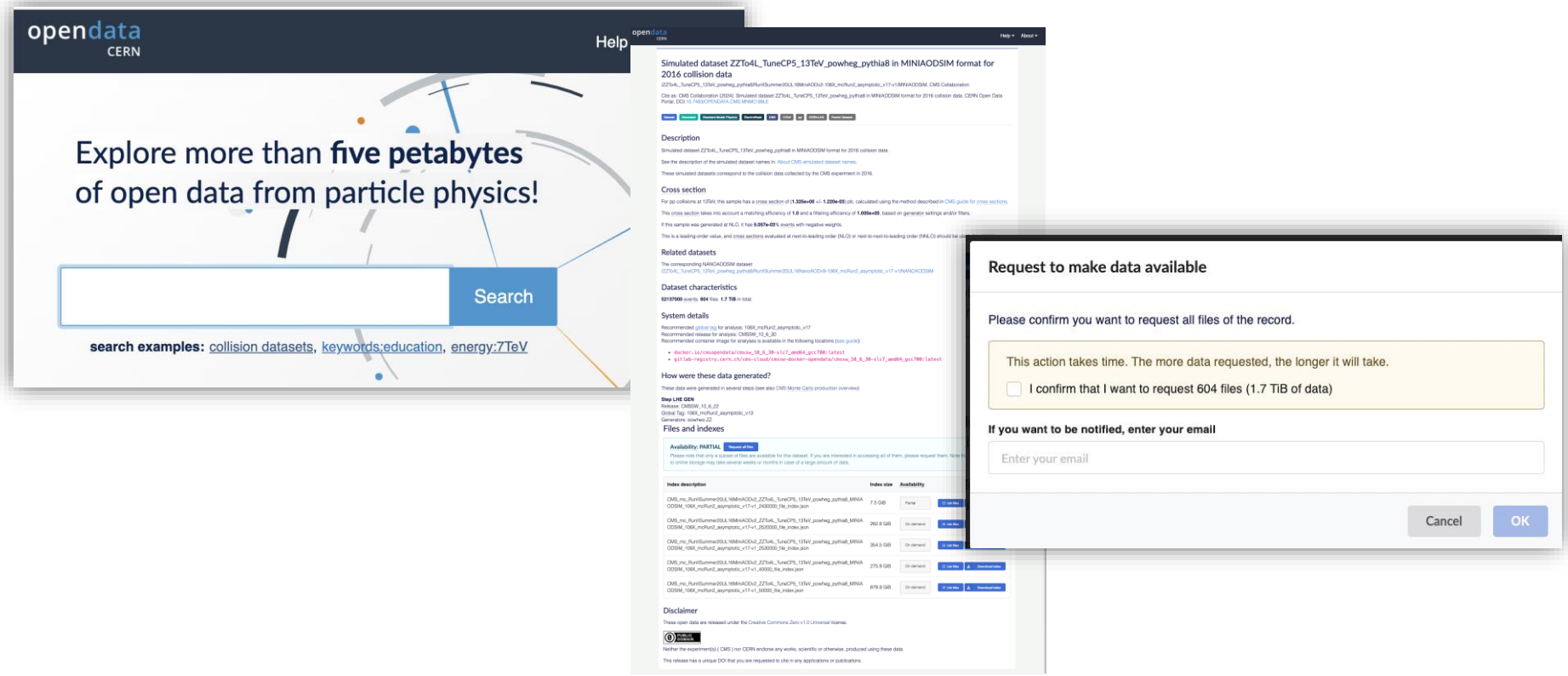
```
[ ]
lumi = 36.6 # fb-1 # data size of the full release
fraction = 1.0 # reduce this is if you want the code to run quicker
```



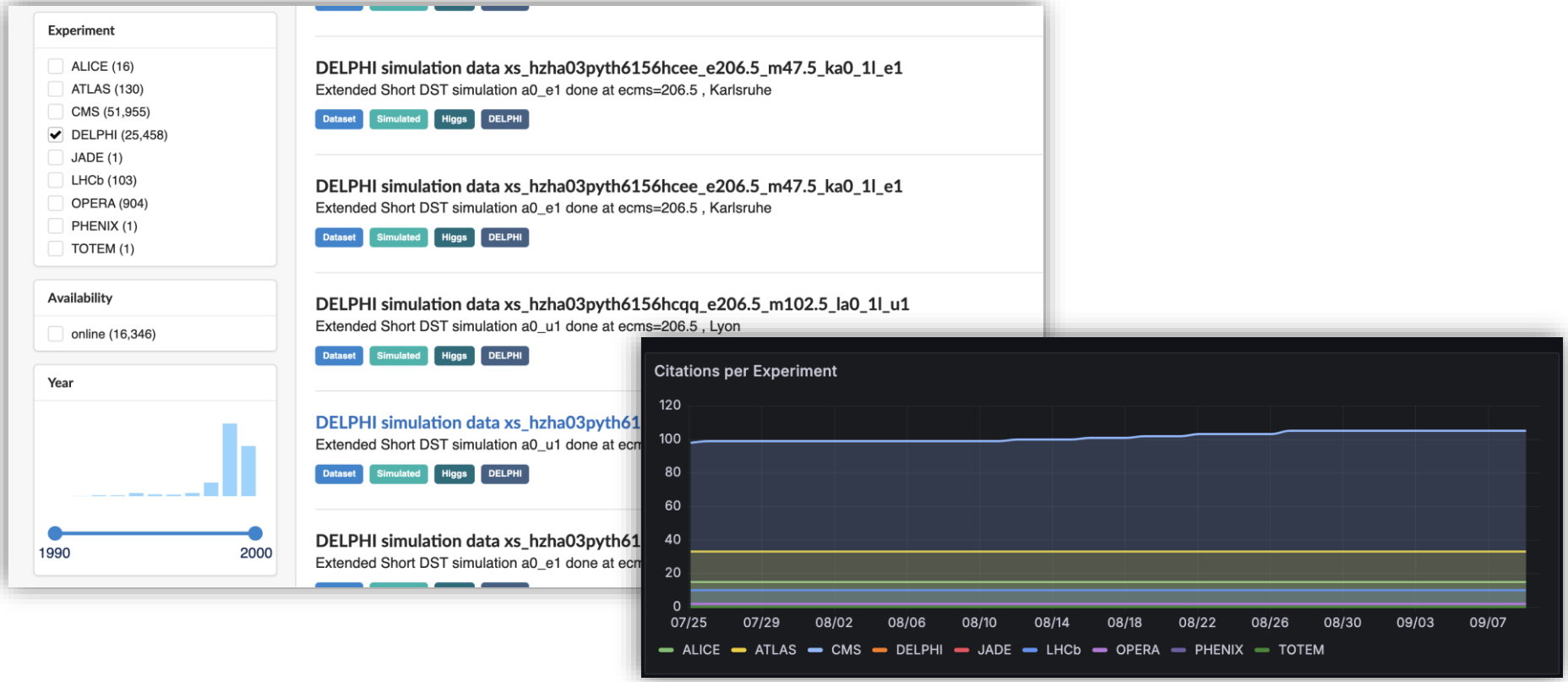


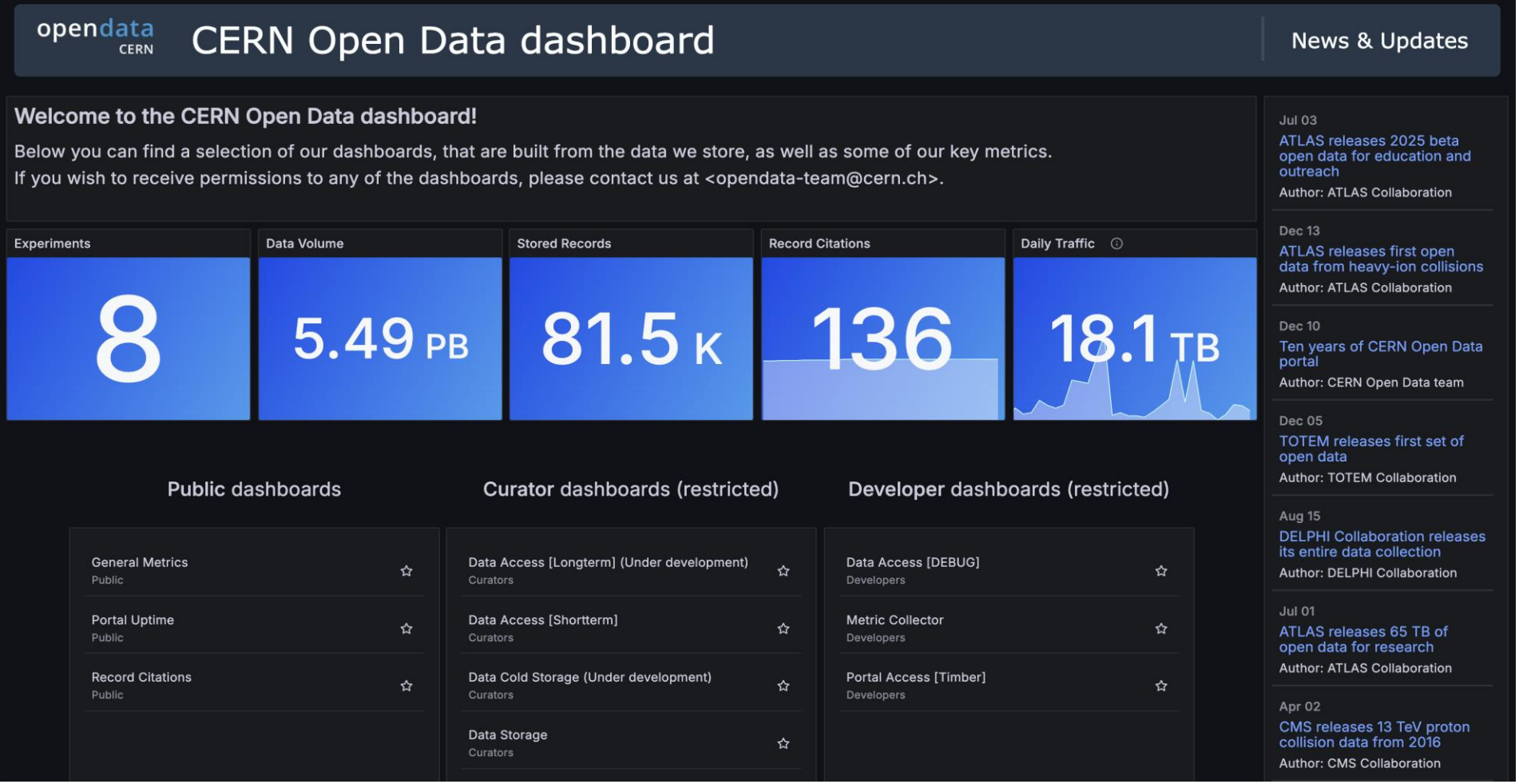
This cartoon is part of a series of post cards raising awareness for the work of the NFDI4BIOIMAGE consortium. Published under a CC BY 4.0 licence.

Cold Storage for CERN Open Data



Long-term preservation




opendata CERN
CERN Open Data dashboard
News & Updates
Welcome to the CERN Open Data dashboard!
Below you can find a selection of our dashboards, that are built from the data we store, as well as some of our key metrics.
If you wish to receive permissions to any of the dashboards, please contact us at <opendata-team@cern.ch>.
Experiments
8
Data Volume
5.49 PB
Stored Records
81.5 K
Record Citations
136
Daily Traffic
18.1 TB
Public dashboards
General Metrics
Public
Portal Uptime
Public
Record Citations
Public
Curator dashboards (restricted)
Data Access [Longterm] (Under development)
Curators
Data Access [Shortterm]
Curators
Data Cold Storage (Under development)
Curators
Data Storage
Curators
Developer dashboards (restricted)
Data Access [DEBUG]
Developers
Metric Collector
Developers
Portal Access [Timber]
Developers
Jul 03
ATLAS releases 2025 beta open data for education and outreach
Author: ATLAS Collaboration
Dec 13
ATLAS releases first open data from heavy-ion collisions
Author: ATLAS Collaboration
Dec 10
Ten years of CERN Open Data portal
Author: CERN Open Data team
Dec 05
TOTEM releases first set of open data
Author: TOTEM Collaboration
Aug 15
DELPHI Collaboration releases its entire data collection
Author: DELPHI Collaboration
Jul 01
ATLAS releases 65 TB of open data for research
Author: ATLAS Collaboration
Apr 02
CMS releases 13 TeV proton collision data from 2016
Author: CMS Collaboration


Home › Dashboards › Developers › Portal Access [Timber]
Search or jump to...
Add
Share
Last 30 days
Environment
prod
Metric Intervals
auto
HTTP Status
All
Request Category
All
Pod Name
All
Filter Requestor
Bot and User
IP Filter
Disabled
Remote IP
0.0.0.0/0
Path
.*
Filters
Raw Access Logs (1 panel)
Numbers
Request success
99.9%
Handled requests
3.03 Mil
Portal traffic
16.9 TB
Downloaded files
36.6 K
Average Response Time
52.7 ms
Distinct countries
156
Distinct Count
Nginx status (2 panels)
Instance requests (6 panels)
Response Times (3 panels)
Paths and IPs (6 panels)
Users
Top browser share
100%
80%
60%
40%
20%
0%
08/20 08/22 08/24 08/26 08/28 08/30 09/01 09/03 09/05 09/07 09/09 09/11 09/13 09/15 09/17
Edge bingbot Chrome misc crawler Firefox Googlebot Safari Python Requests SemrushBot ClaudeBot UT-Dorkbot
UptimeRobot crawler DataForSeoBot GPTBot Opera Amazonbot com/bot Wget Ai2Bot
Top browser share
100%
80%
60%
40%
20%
0%
08/20 08/22 08/24 08/26 08/28 08/30 09/01 09/03 09/05 09/07 09/09 09/11 09/13 09/15 09/17
unknown Windows 10 Mac OSX Android Linux iPhone Windows 7 iPad ChromeOS Windows 8.1 iOS Mac OS X
Windows 2000 Windows XP Other Windows 8 Windows Vista Windows Windows UNKNOWN Ver Xbox One
Top Countries
212455
2063211
United States 71%
Spain 7%
Germany 4%
Switzerland 4%
Singapore 2%
Japan 2%
Brazil 2%
United Kingdom 1%
China 1%
Russia 1%
France 1%
India 1%
Vietnam 1%
Hong Kong 1%
Request locations

CERN Open Data Portal

- **Publicly available repository with more than 5 PB of physics data (and growing!)**
  - 8 experiments, 80.000 records, 3.000.000 files
- **Curated by our experiment colleagues**
  - Datasets, metadata, provenance, tutorials, etc…
- **Following FAIR principles**
  - Findable
  - Accessible
  - Interoperable
  - Reusable
- **Used for education, outreach, independent studies**

home.cern

backup

El modelo estándar

Imágenes:
www.particlezoo.net

MATERIA ORDINARIA

LEPTONS

ELECTRON 

ELECTRON NEUTRINO 

MUON 

MUON NEUTRINO 

TAU 

TAU NEUTRINO 

QUARKS

UP 

DOWN 

CHARM 

STRANGE 

TOP 

BOTTOM 

HIGGS 

4 forces

GLUONS

Fuerza nuclear fuerte

PHOTONS

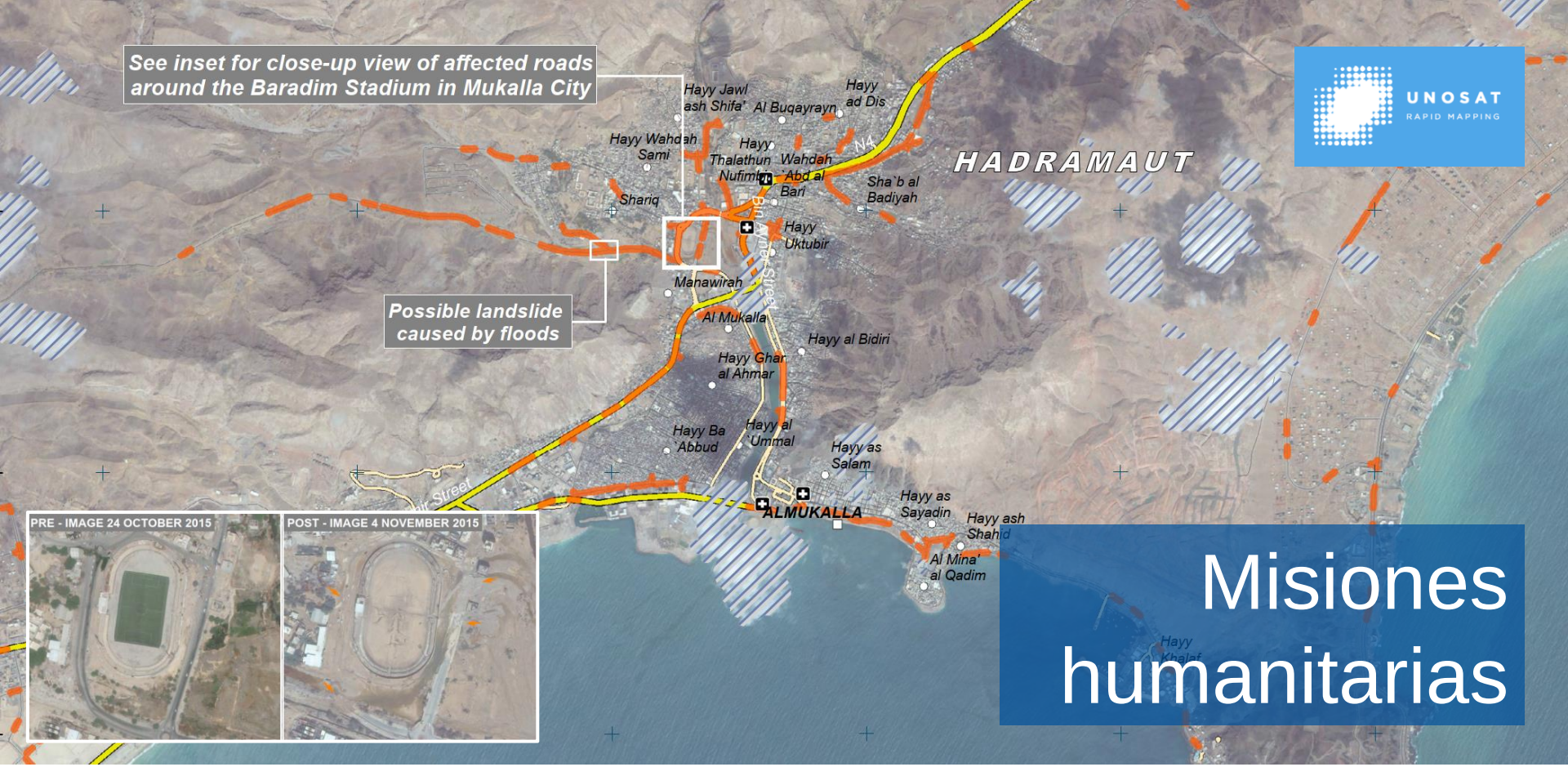
Fuerza electromagnética

BOSONS W&Z

Fuerza nuclear débil

GRAVITON

Gravedad



Misiones humanitarias