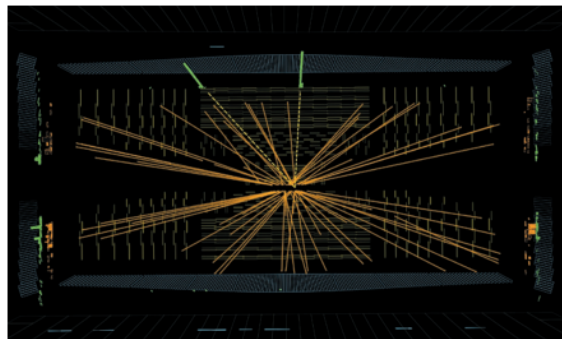




دسترسی به داده‌های آزاد سی.ام.اس

علی فهیم

دانشکده علوم مهندسی - دانشگاه تهران



Data from the Large Hadron Collider, such as this decay of a Higgs boson, could be made publicly available.

PHYSICS

LHC plans for open data future

Researchers share results to keep them accessible.

BY ELIZABETH GIBNEY

When the Large Hadron Collider (LHC) is humming along, the data come in a deluge. The four experimental detectors at the facility, based at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, collect some 25 petabytes of information each year.

Storing the data is not a problem: hard drives are cheap and getting cheaper. The challenge is preserving knowledge that is less commonly stored — the software, algorithms and reference plots specific to each experiment. These often degrade or disappear with time, says Cristinel Diaconu of the Marseilles Centre for Particle Physics in France, who is chair of the international Data Preservation in Long Term Analysis in High Energy Physics (DPHEP) study group. He worries that if the data continue to be stored in their current state, physicists trying to decipher them in 10 years' time will be unable to reconstruct the discovery of the Higgs boson. "When the LHC programme comes to an end, it will probably be the last data at this frontier for many years," he says. "We can't afford to lose it."

The DPHEP is therefore trying to push data-preservation efforts from mere storage to a system of open sharing. The thinking goes that data and the knowledge needed to interpret them are more likely to survive in the long

term if many people outside an experiment are constantly trying to make sense of them.

Kati Lassila-Perini, a physicist at the Compact Muon Solenoid (CMS), one of the four experiments at the LHC, has a radical idea for this sort of sharing: giving data away to school pupils. Next year, a pilot scheme she leads will release 2010 CMS data, which the IT Center for Science in Espoo, Finland, will reformat and store. The centre will then share the data with pupils, who will recreate plots of particle decays using analysis tools adapted for the public. The CMS plans to make more data publicly available a few years after collection, and Lassila-Perini hopes that other data centres will adopt such schemes. "We are guaranteeing that the data we are not looking at any more remain accessible," she says.

The intent is not just to keep data for posterity. Old data can be mined to test new theories and provide crucial references for new experiments, says Diaconu. Before the Higgs boson was discovered in 2012, for example, the Large Electron-Positron collider — the LHC's predecessor at CERN — came back into the spotlight as physicists scoured its 1990s-era data, looking for an exotic type of Higgs that had not been theorized at the time the data had been gathered. In this way, the goals of keeping data alive and open are "enlightened self-interest", says Michael Hildreth, a physicist at

the University of Notre Dame in Indiana and leader of the US-funded Data and Software Preservation for Open Science (DASPOS) effort, which has similar goals to the DPHEP.

DASPOS is building a template for preserving data — a checklist of items that should be stored, and how to do it. Next year, in a 'curation challenge', DASPOS will task physicists with recreating results from other experiments using only the information collected with this template. One test will almost certainly use LHC data — challenging, for example, CMS physicists to recreate results from the rival ATLAS experiment. Another test could come from a different field, such as astrophysics. If successful, the model could form a generic and simplified architecture for preserving data, says Hildreth.

Part of the challenge is coping with ever-changing algorithms, operating systems and data-analysis hardware. At the German Electron Synchrotron (DESY) in Hamburg, computing coordinator David South is leading a project that is already attempting to protect data in this way. His team has devised a system that will automatically comb through data and software from experiments on DESY's Hadron-Electron Ring Accelerator and test them for compatibility when hardware or operating systems change.

This plan to migrate data repeatedly onto new platforms stands in contrast to an approach at the BaBar experiment at the SLAC National Accelerator Laboratory in Menlo Park, California. There, versions of data and the operating systems needed to analyse them have been frozen in storage centres, where they are supposed to be accessible until at least 2018. South says that DESY's approach is more reliable. Although DESY's system needs monitoring — any incompatibilities must be fixed through human intervention — the goal is to deal with problems as they arise, rather than tackle them years later, when they may have compounded.

"When the LHC programme comes to an end, it will probably be the last data at this frontier for many years. We can't afford to lose it."

DESY scientists would know about that. In the 1990s, physicists wanted to take another look at data from a DESY collider that ran from 1979 to 1986, to further investigate the strong interaction that binds quarks together. They managed to measure it with increased precision, but Diaconu says that it took two years to reconstruct the data, which had not been maintained.

The data preservationists are quick to point out the expense associated with reconstruction efforts. Of course, preservation also costs money, but it is well worth it, says DPHEP project manager Jamie Shiers. He puts the bill for implementing good data-preservation at the LHC at around 1% of operating costs — just a few million dollars per year. "I think it's justified," he says. ■

چشم انداز

- حفظ داده ها و دانش (فراداده)
- اشتراک گذاری آزاد. داده ها و دانش به احتمال زیاد زنده می ماند.
- داده ها را در اختیار دانشجویان و محققان قرار دهید و به آنها اجازه دهید به عنوان مثال برای بازسازی کشف هیگز، به فیزیکدانان CMS اجازه دهید نتایج ATLAS را بازسازی کنند و بالعکس.
- استخراج داده ها برای آزمایش نظریه های جدید و ارائه منابع مهم
- هزینه تقریباً یک درصد از هزینه های عملیاتی است پس ارزش تلاش را دارد.

محتوای سامانه‌ی داده‌های آزاد سرن

در نوامبر 2014 راه اندازی شد.

محتوای بسیار غنی شامل:

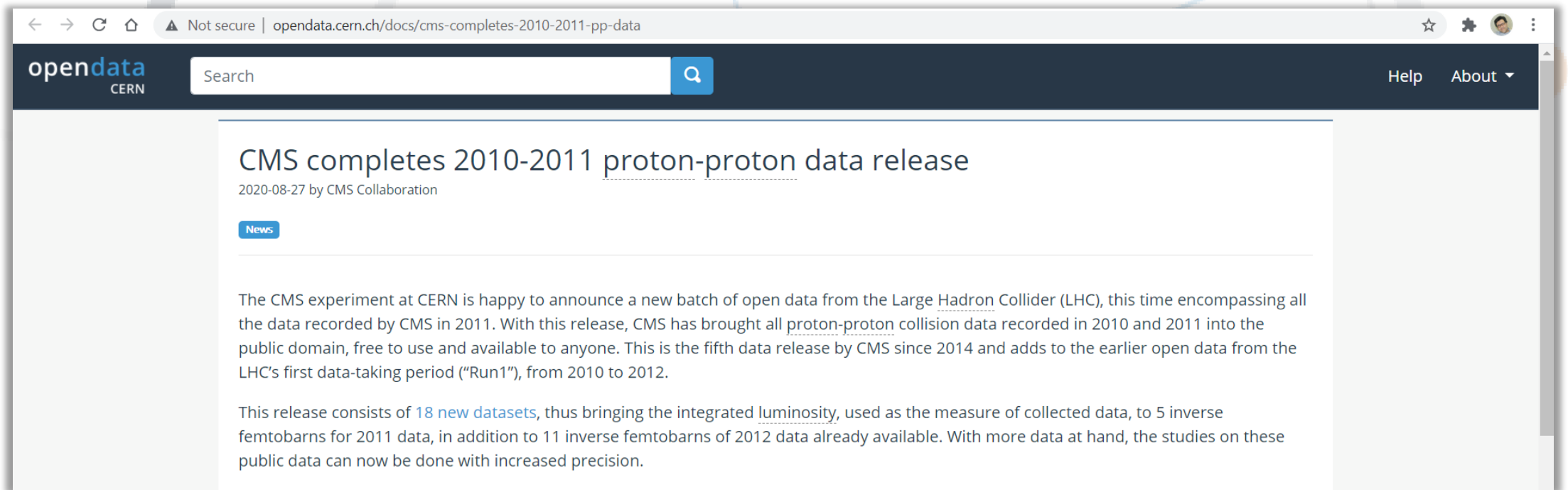
- مجموعه داده‌های برخوردی واقعی و شبیه‌سازی شده قابل استفاده برای پژوهش
- مجموعه داده‌های آماده شده برای آموزش دانشجویان
- فایل‌های پیکربندی (Configuration files) و مستندات
- ماشین‌های مجازی و تصاویر محتوا (Container images)
- ابزارهای نرم افزاری و نمونه‌های تجزیه و تحلیل

بیش از 2 پتابایت

توسعه یافته توسط CERN-IT و همکاری آزمایشگاه‌های:



محتوای سامانه‌ی داده‌های آزاد سرن



The screenshot shows a web browser window with the URL `opendata.cern.ch/docs/cms-completes-2010-2011-pp-data`. The page header includes the "opendata CERN" logo, a search bar, and navigation links for "Help" and "About". The main content area features a news article titled "CMS completes 2010-2011 proton-proton data release", dated "2020-08-27 by CMS Collaboration". A blue "News" tag is positioned above the article text. The article text states: "The CMS experiment at CERN is happy to announce a new batch of open data from the Large Hadron Collider (LHC), this time encompassing all the data recorded by CMS in 2011. With this release, CMS has brought all proton-proton collision data recorded in 2010 and 2011 into the public domain, free to use and available to anyone. This is the fifth data release by CMS since 2014 and adds to the earlier open data from the LHC's first data-taking period ("Run1"), from 2010 to 2012." A second paragraph follows: "This release consists of 18 new datasets, thus bringing the integrated luminosity, used as the measure of collected data, to 5 inverse femtobarns for 2011 data, in addition to 11 inverse femtobarns of 2012 data already available. With more data at hand, the studies on these public data can now be done with increased precision."

بستر داده‌های باز

<https://inveniosoftware.org/showcase/>

INVENIO
Powering Open Science

Framework
Open Source framework for large-scale digital repositories

RDM
Turn-key Research Data Management repository

ILS
Integrated Library System

Examples Community Docs Blog Talk About

Instances. Add yours!

All Data Library Multimedia Repository

CERN Document Server
CERN official repository for publications, articles, reports and multimedia content in HEP.
Library Multimedia Repository

CERN Open Data
Discover open Research datasets and software of LHC experiments. Visualise events and run your own analysis.
Data

HEPData
The Durham High Energy Physics Database (HEPData) has been built up over the past four decades as a unique open-access repository for scattering data from experimental particle physics. It currently comprises the data points from plots and

پُرتال داده‌های آزاد سرن

http://opendata.cern.ch/

The screenshot shows the OpenData CERN website. At the top, there is a navigation bar with the 'opendata CERN' logo on the left and 'Help' and 'About' links on the right. The main content area features a large heading: 'Explore more than two petabytes of open data from particle physics!'. Below this is a search bar with the placeholder text 'Start typing...' and a blue 'Search' button. Underneath the search bar, there are search examples: 'collision datasets', 'keywords:education', and 'energy:7TeV'. The page is divided into two columns. The left column is titled 'Explore' and contains links for 'datasets', 'software', 'environments', and 'documentation'. The right column is titled 'Focus on' and lists several particle physics experiments: 'ATLAS', 'ALICE', 'CMS', 'LHCb', 'OPERA', 'PHENIX', and 'Data Science'. The background of the website has a decorative pattern of overlapping circles and lines in shades of blue and orange.

پرتال داده‌های آزاد سرن

search examples: collision datasets

The screenshot shows the OpenData CERN search interface. The browser address bar displays the URL: `opendata.cern.ch/search?page=1&size=20&subtype=Collision&type=Dataset`. The page header includes the OpenData CERN logo, a search bar, and navigation links for Help and About. The main content area is divided into a left sidebar for filtering and a right section for search results.

Filtering: The left sidebar shows filters for Dataset (2202) and Collision (163). Other filters include Documentation (67), Environment (38), and Software (46).

Search Results: The results are sorted by Best match (ascending) and displayed in a detailed view. Three results are visible:

- /ZeroBias/Commissioning10-May19Reco-v1/RECO**: ZeroBias primary dataset in RECO format from the 0.9 and 7 TeV Commissioning runs of 2010. This dataset includes the data from the CASTOR calorimeter...
- /MinBias0Tesla1/Run2011A-PromptReco-v5/RECO**: MinBias0Tesla1 primary dataset from the 7 TeV proton-proton run of 2011. These proton-proton data are at the same centre-of-mass energy and have a similar trigger menu to those in Pb-Pb collisions....
- /SingleElectron/Run2012B-v1/RAW**: A sample from SingleElectron primary dataset in RAW format from RunB of 2012. Run range [194117, 194199]. This dataset contains selected runs from 2012 RunB. The list of validated lumi secti...

پُرتال داده‌های آزاد سرن

The screenshot shows a web browser window with the URL `opendata.cern.ch/#lva`. The page features a navigation bar with a "Get started" dropdown menu. Below the navigation bar, there are three main sections: "Learn", "Visualise", and "Analyse". Each section has a brief description and a list of links. The "Learn" section includes links for "Welcome to our updated portal", "CMS Guide to education use of CMS Open Data", "Improving educational content with high school teachers: A field report from our summer students", and "Glossary". The "Visualise" section includes links for "CMS Event Display", "OPERA Event Display", and "CMS Histograms". The "Analyse" section includes links for "CMS Guide to research use of CMS Open Data", "ATLAS Higgs Machine Learning Challenge", "Getting Started with LHCb Open Data", and "Getting Started with ALICE Open Data". Each section also has a "more" button at the bottom.

← → ↻ 🏠 ⚠ Not secure | opendata.cern.ch/#lva ☆ ⚙ 👤 ⋮

∨ Get started ∨

Learn

Discover the world of open data from particle physics

Welcome to our updated portal
CMS Guide to education use of CMS Open Data
Improving educational content with high school teachers: A field report from our summer students
Glossary

more

Visualise

Explore detector events and run basic histogramming

CMS Event Display
OPERA Event Display
CMS Histograms

Analyse

Run your own physics analyses, start virtual machines

CMS Guide to research use of CMS Open Data
ATLAS Higgs Machine Learning Challenge
Getting Started with LHCb Open Data
Getting Started with ALICE Open Data

more

Education



The CMS (Compact Muon Solenoid) experiment is one of two large general-purpose detectors built on the Large Hadron Collider (LHC). Its goal is to investigate a wide range of physics such as the characteristics of the Higgs boson, extra dimensions or dark matter.

[Explore CMS >](#)



ALICE

ALICE (A Large Ion Collider Experiment) is a heavy-ion detector designed to study the physics of strongly interacting matter at extreme energy densities, where a phase of matter called quark-gluon plasma forms. More than 1000 scientists are part of the collaboration.

[Explore ALICE >](#)



The ATLAS (A Toroidal LHC ApparatuS) experiment is a general purpose detector exploring topics like the properties of the Higgs-like particle, extra dimensions of space, unification of fundamental forces, and evidence for dark matter candidates in the Universe.

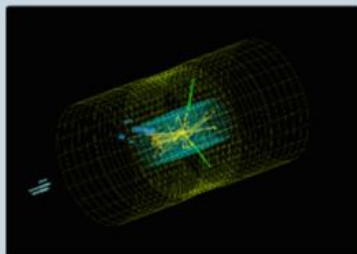
[Explore ATLAS >](#)



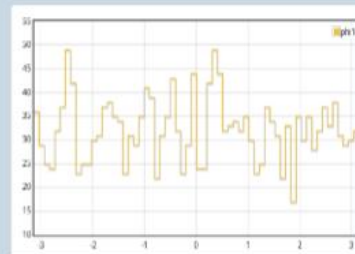
The LHCb (Large Hadron Collider beauty) experiment aims to record the decay of particles containing b and anti-b quarks, known as B mesons. The detector is designed to gather information about the identity, trajectory, momentum and energy of each particle.

[Explore LHCb >](#)

For education purposes, the complex primary data need to be processed into a format (examples below) that is good for simple applications. Get in touch if you wish to build your own applications similar to those shown here



[Visualise events >](#)



[Visualise histograms >](#)



[Learning Resources >](#)

انتشار مقالات

PHYSICAL REVIEW D
covering particles, fields, gravitation, and cosmology

Highlights Recent Accepted Collections Authors Referees Search Press About Staff

Editors' Suggestion

Jet substructure studies with CMS open data

Aashish Tripathy, Wei Xue, Andrew Larkoski, Simone Marzani, and Jesse Thaler
Phys. Rev. D **96**, 074003 – Published 3 October 2017

14

Twitter Facebook More

Article References Citing Articles (22) PDF HTML Export Citation

ABSTRACT

We use public data from the CMS experiment to study the two-prong substructure of jets. The CMS open data are based on 31.8 pb^{-1} of 7 TeV proton-proton collisions recorded at the Large Hadron Collider in 2010, yielding a sample of 768,687 events containing a high-quality central jet with transverse momentum larger than 85 GeV. Using CMS's particle flow reconstruction algorithm to obtain jet constituents, we extract the two-prong substructure of the leading jet using soft-drop declustering. We find good agreement between results obtained from the CMS open data and those obtained from parton shower generators, and we also compare to analytic jet substructure calculations performed to modified leading-logarithmic accuracy. Although the 2010 CMS open data do not include simulated data to help estimate systematic uncertainties, we use track-only observables to validate these substructure studies.

Issue
Vol. 96, Iss. 7 — 1 October 2017

Reuse & Permissions

Access Options
[Buy Article »](#)

- از افتخارات این پرتال این است که داده‌های قدیمی نتایج جدیدی را به ارمغان آورده است.
- در سال ۲۰۱۷ اولین مقاله علمی با استفاده از داده‌های LHC توسط محققان خارج از آزمایشگاه منتشر شد.

← → ↻ 🏠 scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q="cms+open+data"&btnG=

☰ Google Scholar "cms open data" 🔍

📌 Articles About 251 results (0.03 sec)

- Any time
- Since 2021
 - Since 2020
 - Since 2017
 - Custom range...
- Sort by relevance
- Sort by date
- include patents
- include citations
- Create alert

- Exploring the space of jets with CMS open data** [PDF] aps.org
[PT Komiske](#), [R Mastandrea](#), [EM Metodiev](#), [P Naik](#)... - Physical Review D, 2020 - APS
 We explore the metric space of jets using public collider data from the CMS experiment. Starting from 2.3 fb⁻¹ of proton-proton collisions at s = 7 TeV collected at the Large Hadron Collider in 2011, we isolate a sample of 1,690,984 central jets with transverse momentum ...
 ☆ 📄 Cited by 14 Related articles All 9 versions
- Jet substructure studies with CMS open data**
[A Tripathee](#), [W Xue](#), [A Larkoski](#), [S Marzani](#), [J Thaler](#) - Physical Review D, 2017
 We use public data from the CMS experiment to study the two-prong substructure. CMS open data are based on 31.8 pb⁻¹ of 7 TeV proton-proton collisions recorded at the Large Hadron Collider in 2010, yielding a sample of 768,687 events containing ...
 ☆ 📄 Cited by 53 Related articles All 8 versions
- Exposing the QCD splitting function with CMS open data**
[A Larkoski](#), [S Marzani](#), [J Thaler](#), [A Tripathee](#), [W Xue](#) - Physical review letters, 2019
 The splitting function is a universal property of quantum chromodynamics (QCD) that describes how energy is shared between partons. Despite its ubiquitous appearance in many QCD calculations, the splitting function cannot be measured directly, since ...
 ☆ 📄 Cited by 72 Related articles All 10 versions
- Adversarially Learned Anomaly Detection on CMS Open Data: discovering the top quark**
[O Knapp](#), [O Cerri](#), [G Dissertori](#), [TQ Nguyen](#)... - The European Physical Journal Special Topics, 2021
 Abstract We apply an Adversarially Learned Anomaly Detection (ALAD) algorithm to the problem of detecting new physics processes in proton-proton collisions at the LHC. Anomaly detection based on ALAD matches performances reached by other methods.
 ☆ 📄 Cited by 22 Related articles All 10 versions
- End-to-end jet classification of quarks and gluons with deep learning**
[M Andrews](#), [J Alison](#), [S An](#), [B Burkle](#), [S Gleyzer](#)... - Nuclear Instruments and Methods in Physics Research Section A, 2021
 We describe the construction of novel end-to-end jet image classifiers to discriminate between quark-initiated jets and gluon-initiated jets using the simulated CMS Open Data. These multi-detector classifiers correspond to true maps of the low-level energy deposits in the detector, giving ...

انتشار مقالات


← → ↻ 🏠 iopscience.iop.org/article/10.1088/1361-6471/ab33a9

IOPscience 🔍 Journals ▾ Books Publishing Support 🌐 Login ▾

Journal of Physics G: Nuclear and Particle Physics

PAPER

Explicit jet veto as a tool to purify the underlying event in the Drell–Yan process using CMS Open Data



Saeid Paktinat Mehdiabadi^{4,1,2}  and Ali Fahim³

Published 9 August 2019 • © 2019 IOP Publishing Ltd

[Journal of Physics G: Nuclear and Particle Physics, Volume 46, Number 9](#)

Citation Saeid Paktinat Mehdiabadi and Ali Fahim 2019 *J. Phys. G: Nucl. Part. Phys.* **46** 095003





23 Total downloads

[Turn on MathJax](#)

[Get permission to re-use this article](#)

Share this article

[+ Article information](#)

Abstract

The underlying event (UE) is an important part of high-energy collision events. In the event generators, the UE is tuned by fits to collision data. Usually, the UE observables are affected by the existence of extra jets and it is difficult to find a part of the phase space which is dominated by the UE. In this paper, we suggest to veto the jets in the considered region to disentangle these effects. The idea is verified to work on CMS Open Data. To our knowledge, it is the first time that such ideas are tested on real collision data.

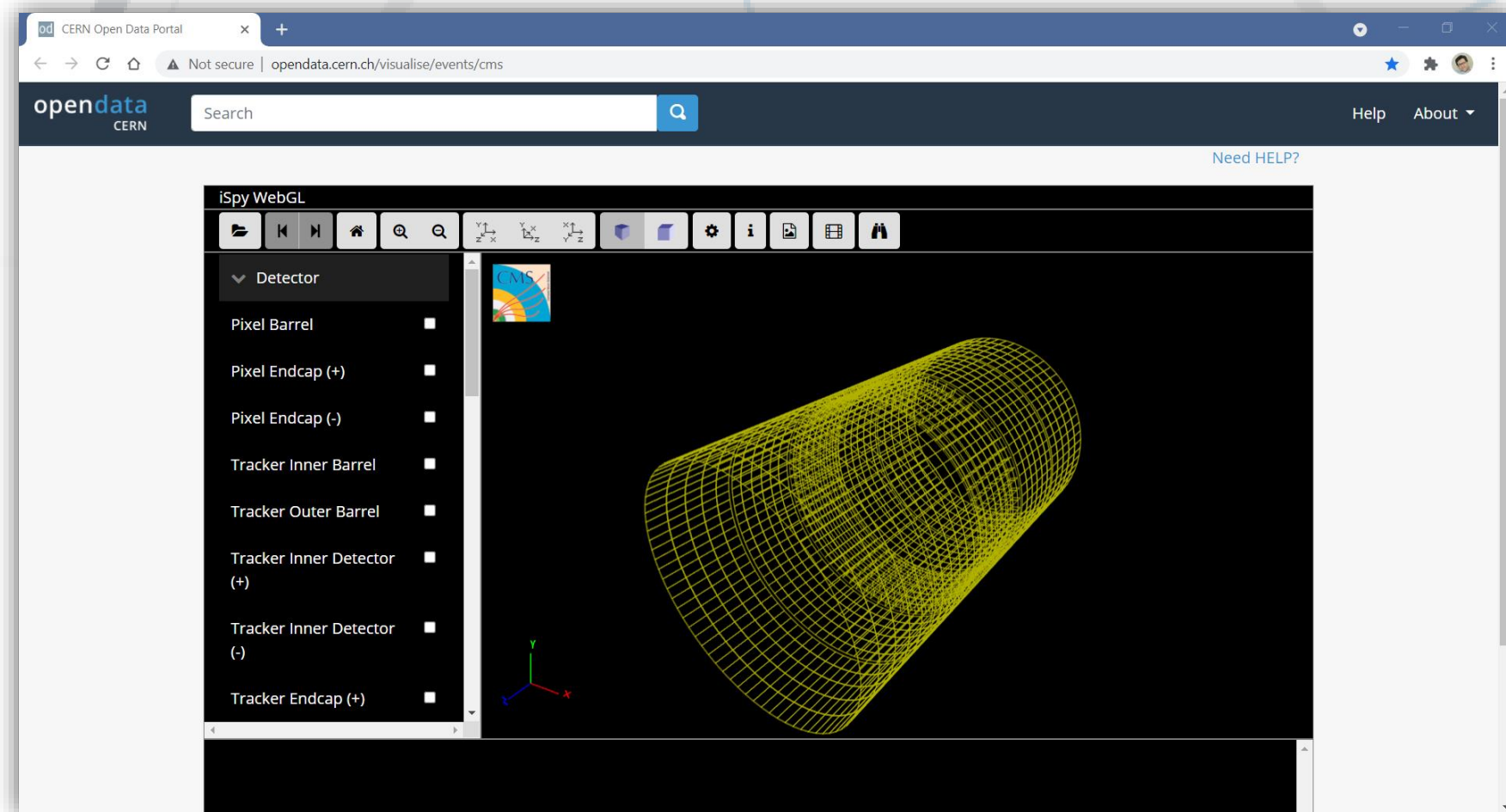
Abstract

راهنمای استفاده از داده‌های آزاد CMS برای آموزش

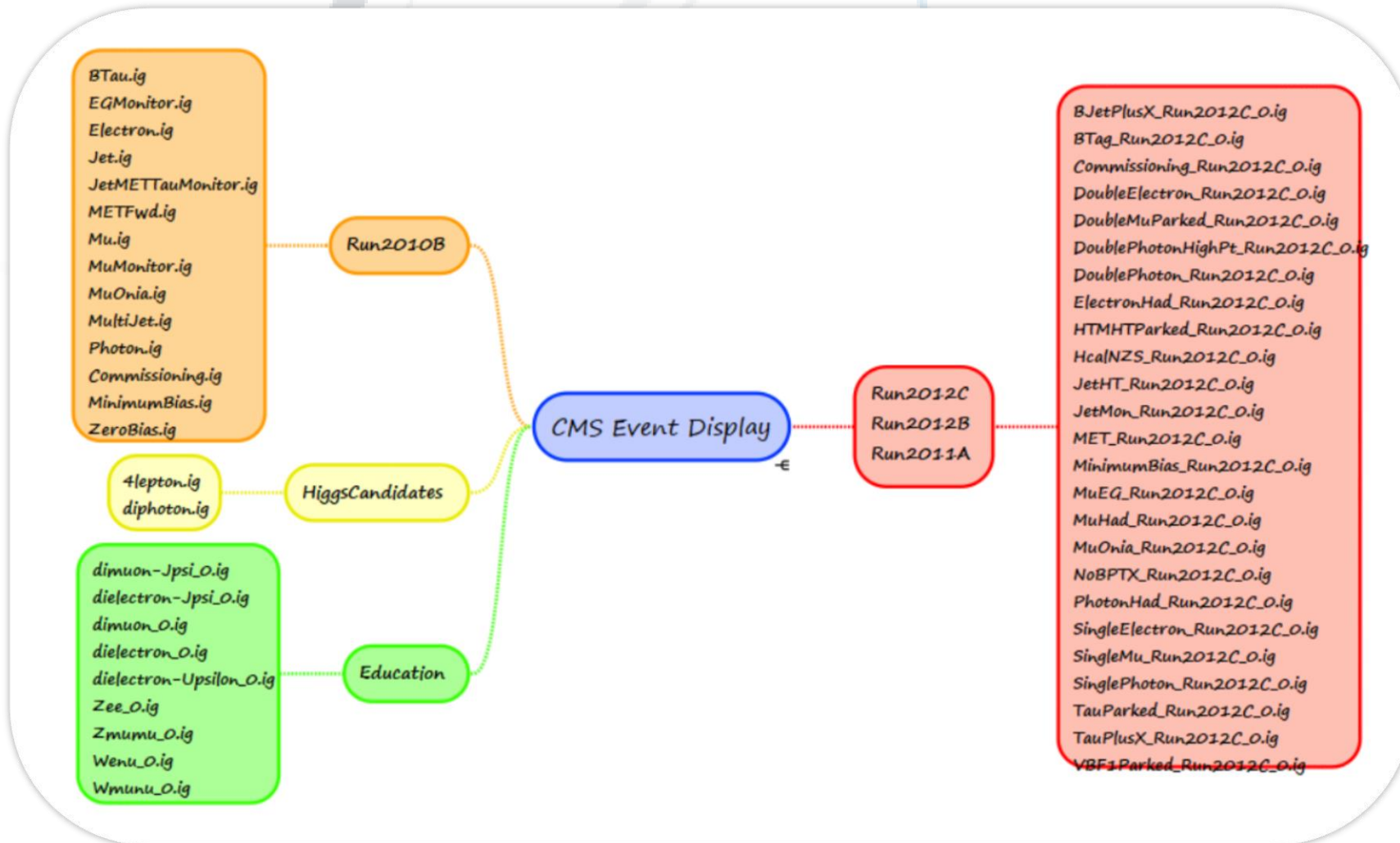
- نمایش برخورد ذرات
- تهیه هیستوگرام با استفاده از داده‌های برخورد ذرات
- ورود عمیق‌تر در داده‌های برخورد

نمایش برخورد ذرات

<http://opendata.cern.ch/visualise/events/cms>



نمایش برخوردار ذرات



نمایش برخورد ذرات

HiggsCandidates/4lepton.ig/Run_193575/Event_400912970

The screenshot displays the opendata.cern.ch website interface. The browser address bar shows the URL: `opendata.cern.ch/visualise/events/cms`. The page title is "4lepton.ig:Events/Run_193575/Event_400912970 [2 of 3]". The main content area features a 3D visualization of particle tracks and calorimeter hits, with a coordinate system (x, y, z) visible. A sidebar on the left lists detector components with checkboxes:

- ECAL Endcap (+)
- ECAL Endcap (-)
- HCAL Barrel
- HCAL Endcap (+)
- HCAL Endcap (-)
- HCAL Outer
- HCAL Forward (+)
- HCAL Forward (-)
- Drift Tubes
- Cathode Strip Chambers

Metadata information is displayed in the top right of the visualization area:

- CMS Experiment at the LHC, CERN
- Data recorded: 2012-May-07 07:46:20.384985 GMT
- Run / Event / LS: 193575 / 400912970 / 523

A footer note reads: "Click on a name under 'Provenance', 'Tracking', 'ECAL', 'HCAL', 'Muon', and 'Physics' to view contents in table".

رویدادهای نامزد هیگز

DOI:10.7483/OPENDATA.CMS.N9MJ.QEEC
<http://opendata.cern.ch/record/300>

این سند شامل رویدادهای نامزد هیگز با جرم ثابت بین 120-130 GeV است که توسط CMS برای استفاده در آموزش و توسعه منتشر شده است:

- 10 رویداد گاما گاما،
- 1 رویداد $e2\mu2$ ،
- 1 رویداد $\mu4$
- 1 رویداد $e4$

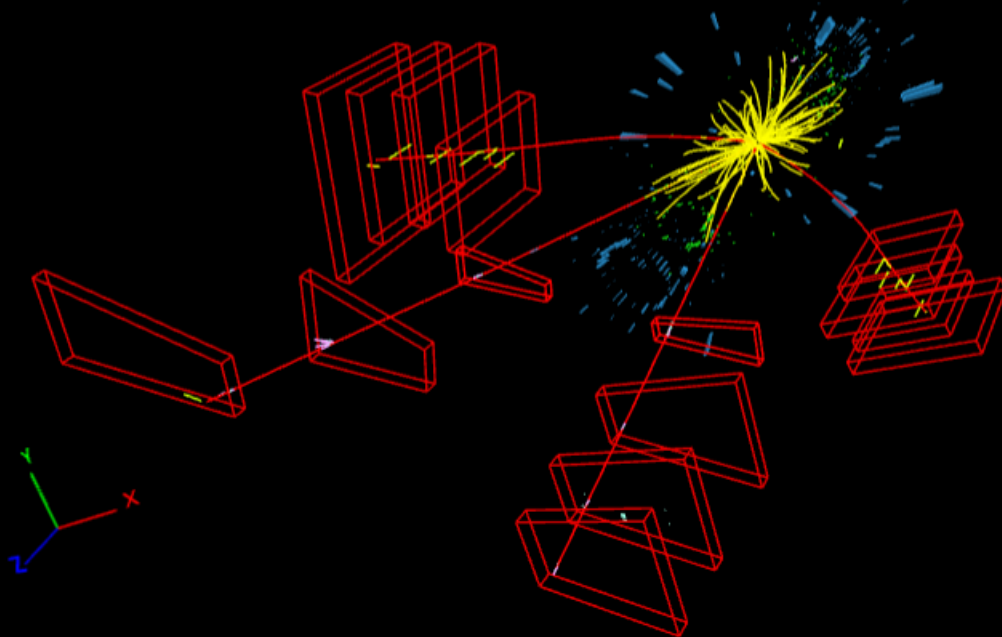
این داده‌ها برای استفاده در آموزش و یادگیری انتخاب شده‌اند و حاوی زیرمجموعه‌ای از کل اطلاعات رویداد هستند. آنها برای تجزیه و تحلیل کامل فیزیک مناسب نیستند.



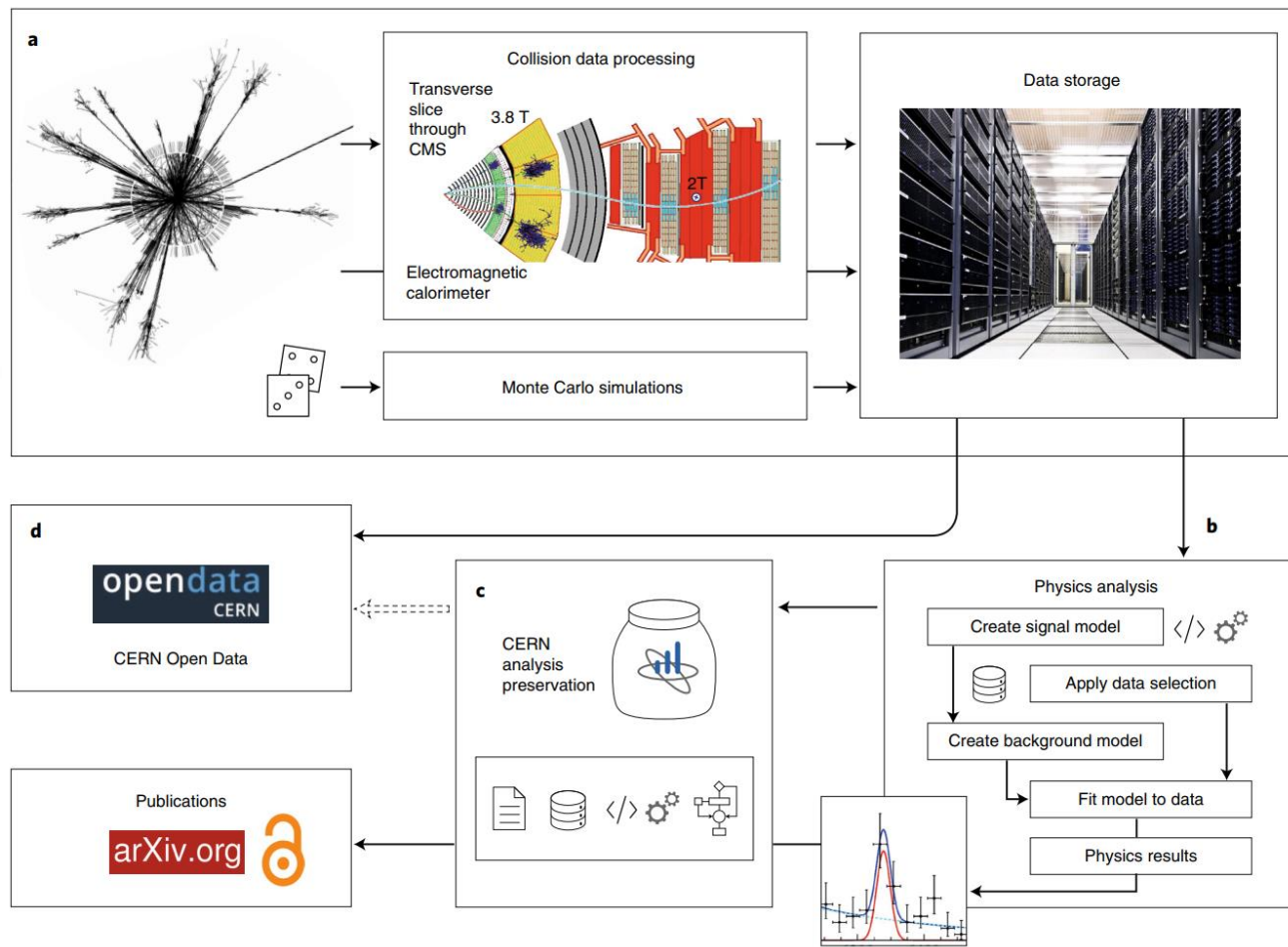
CMS Experiment at the LHC, CERN

Data recorded: 2011-Oct-13 12:47:38.421105 GMT

Run / Event / LS: 178424 / 666626491 / 585



از برخورد تا داده‌ی آزاد



استفاده از ماشین مجازی



virtualbox.org/wiki/Downloads

VirtualBox

Download VirtualBox

Here you will find links to VirtualBox binaries and its source code.

VirtualBox binaries

By downloading, you agree to the terms and conditions of the res...

If you're looking for the latest VirtualBox 6.0 packages, see [VirtualBox 6.1](#). Version 6.0 will remain supported until July 2020.

If you're looking for the latest VirtualBox 5.2 packages, see [VirtualBox 5.2](#). Version 5.2 will remain supported until July 2020.

VirtualBox 6.1.26 platform packages

- ⇒ [Windows hosts](#)
- ⇒ [OS X hosts](#)
- ⇒ [Linux distributions](#)
- ⇒ [Solaris hosts](#)
- ⇒ [Solaris 11 IPS hosts](#)

The binaries are released under the terms of the GPL version 2.

See the [changelog](#) for what has changed.

You might want to compare the checksums to verify the integrity

- [SHA256 checksums](#), [MD5 checksums](#)

Note: After upgrading VirtualBox it is recommended to upgrade t...

CMS VM Image, for 2011 and 2012 CMS open data

CMS Collaboration

[Environment](#) [VM](#) [CMS](#) [CERN-LHC](#)

Description

This virtual machine image provides CMS computing environment to be used with the 2011 and 2012 CMS open data. The virtual machine is based on the CernVM ([cernvm.cern.ch](#)) and uses Scientific Linux CERN. The image gets the CMS software (CMSSW) from [/cvmfs/cms.cern.ch](#) and the jobs running on the CMS open data VM read the condition data from [/cvmfs/cms-opendata-condddb.cern.ch](#). Access to the data is through XRootD.

CMS-OpenData-1.5.3.ova is recommended. It has a 40G virtual hard disk and a 20G cvmfs cache, which is large enough for condition data for full [event range](#) for 2012 data (see [the CMS guide to the condition database](#) for further details). It has an embedded CERN Scientific Linux 6 (slc6) shell, where all CMS software specific commands should be executed. Additionally, it has a CERN CentOS 7 (cc7) shell, which can be used in the same session.

The versions earlier than CMS-OpenData-1.5.3.ova are deprecated; please use the latest available version.

For known issues and limitations see

[CMS Virtual Machines - Known Issues and Limitations](#)

Characteristics

5 files. 98.9 MB in total.

<http://opendata.cern.ch/docs/cms-virtual-machine-2011>

<http://opendata.cern.ch/docs/cms-virtual-machine-2010>

استفاده از ماشین مجازی

⚠ Not secure | opendata.cern.ch/record/252

Files

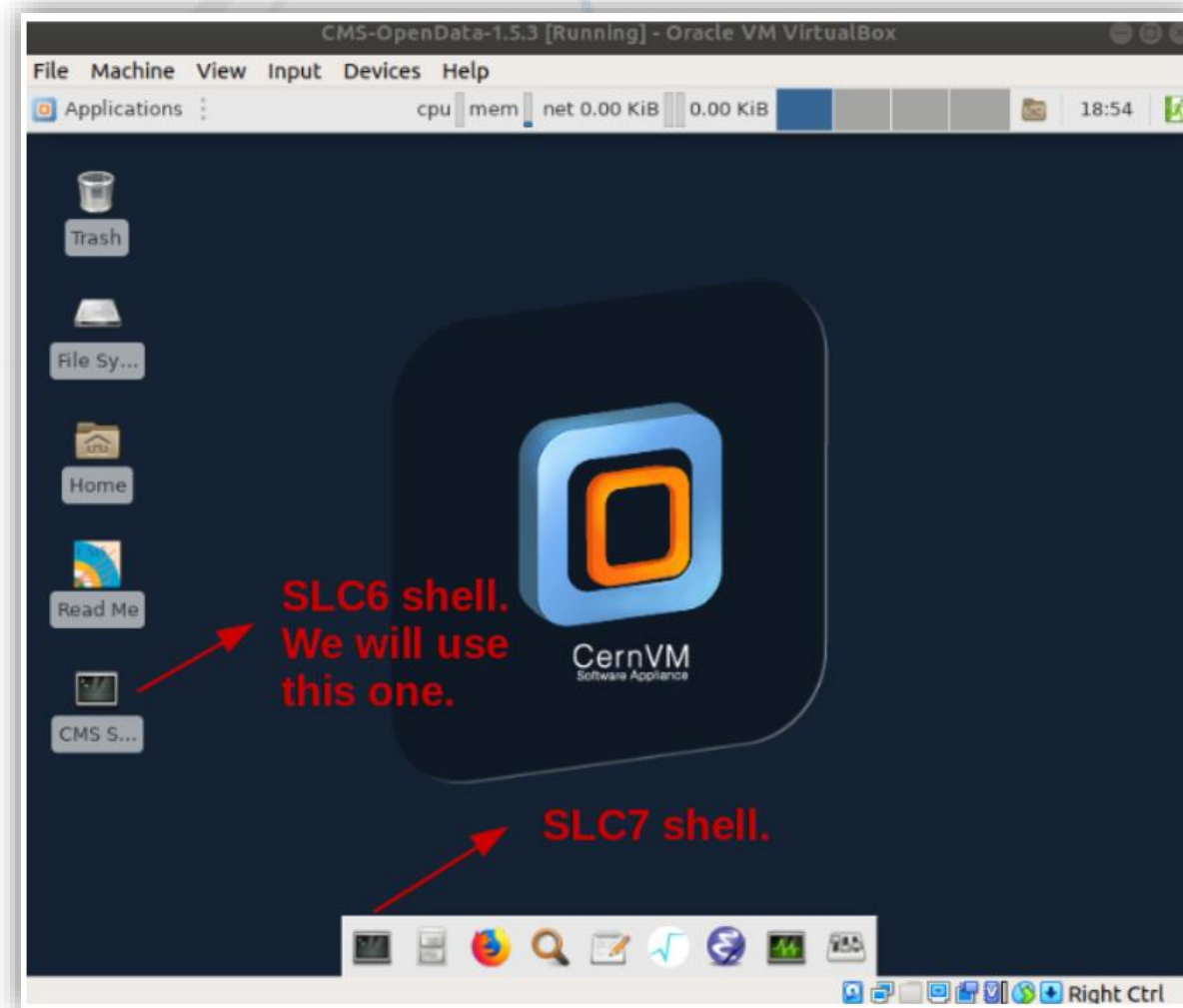
Filename	Size	
CMS-Open-Data-1.2.0.ova	17.3 MB	Download
CMS-Open-Data-1.3.0.ova	17.9 MB	Download
CMS-OpenData-1.5.1.ova	19.7 MB	Download
CMS-OpenData-1.5.2.ova	19.9 MB	Download
CMS-OpenData-1.5.3.ova	19.6 MB	Download

Downloading the CMS-specific CernVM image

Download the 2011 CMS-specific CernVM image as OVA file from the official [record](#) in the Cern Open Portal site. We recommend using version 1.5.3, i.e., [CMS-OpenData-1.5.3](#). This VM Image can be used for data from 2011 and 2012, which are the data we will use in this workshop.

Make sure to download version of the ova file, [1.5.3](#). Other versions are deprecated.

استفاده از ماشین مجازی



استفاده از داکر



These pre-exercises are designed to prepare you for using Docker in the context of the CMS Open Data Workshop. This is *not* a full introduction to Docker, however we do take time to explain some key concepts that should make your experience a smooth one.

Note that you can also use virtual machines (VMs) to interface with the CERN Open Data Portal and this workshop, and there is a separate pre-exercise for that.

You should definitely take time to go through either this *or* the [VM exercise](#) before participating in the workshop.

If you run into problems with any of these steps, please reach out to the organizers through the dedicated [Mattermost channel](#).

استفاده از داکر

Download the docker image for CMS open data and start a container

The first time you start a container, a docker image file gets downloaded from an image registry. The CMS open data image is large and it may take some time to download, even as long as 20-30 minutes, depending on the speed of your internet connection. After the download, a container created from that image starts. The download needs to be done only once. Afterwards, when starting a container, it will find the downloaded image on your computer, and it will be much faster.

Please follow the instructions below, depending on the operating system you are using.

Linux

Windows WSL2

MacOS

Start the image download and open the container with

Bash

```
docker run -it --name my_od -P -p 5901:5901 cmsopendata/cmssw_5_3_32_vnc:latest /bin/bash
```

Output

```
Setting up CMSSW_5_3_32  
CMSSW should now be available.  
~/CMSSW_5_3_32/src $
```

This is now a bash shell in the CMS open data environment in which you have access to a complete CMS software release that is appropriate for interfacing with the 2011 and 2012 7 and 8 TeV datasets.

آدرس داده‌ها

```
opendata.cern.ch/record/6021/files/CMS_Run20...
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/0412B176-AEE3-E211-A700-20CF3019DF03.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/0A2FCA76-AEE3-E211-A8A2-E0CB4E19F9A6.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/0C3C6077-AEE3-E211-A4CA-20CF305B059C.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/0C57EA77-AEE3-E211-8DCA-00259073E4D4.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/0E653900-AEE3-E211-8E25-90E6BA19A266.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/100D9A77-AEE3-E211-BDA9-20CF305B058E.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/1637B176-AEE3-E211-AAA0-90E6BA0D09AF.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/1821BA77-AEE3-E211-A284-002590747DDC.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/189322F2-AEE3-E211-86EF-E0CB4E5536A8.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/1C111592-AEE3-E211-99F8-90E6BA442F33.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/229DA876-AEE3-E211-B91C-90E6BA19A1FE.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/22B85676-AEE3-E211-9D0E-90E6BA19A24F.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/2638AF76-AEE3-E211-A363-E0CB4E19F99B.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/2891BC3B-AEE3-E211-8CAC-BCAEC54B303A.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/2AC9937F-AEE3-E211-BD91-00261834B561.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/32AF576-AEE3-E211-AC1A-00248C9BA537.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/32B05D76-AEE3-E211-8DA1-90E6BA442F00.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/469C1B77-AEE3-E211-902D-20CF3027A5F3.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/46E80177-AEE3-E211-91D7-E0CB4E19F981.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/4A898E77-AEE3-E211-AB0A-0025907277A0.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/4CBEE476-AEE3-E211-B77D-BCAEC53F6D3A.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/50C18777-AEE3-E211-A21F-00259073E488.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/520A4D76-AEE3-E211-B883-90E6BA19A214.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/563E7877-AEE3-E211-9B55-20CF305B04F0.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/620E7977-AEE3-E211-BF0C-90E6BA0D0998.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/62CAAD77-AEE3-E211-A668-00259073E408.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/6A3CA277-AEE3-E211-80EE-E0CB4E5536A5.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/7C9C3077-AEE3-E211-B562-20CF3027A5D5.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/7CA8CA76-AEE3-E211-A353-20CF3019DF03.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/A0A7C576-AEE3-E211-B218-E0CB4E1A1194.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/A46BD076-AEE3-E211-812F-00261834B579.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/AA018F76-AEE3-E211-989C-E0CB4E19F973.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/AA29A376-AEE3-E211-8919-20CF3027A626.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/AC144D76-AEE3-E211-AF26-E0CB4E55363B.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/B66D2A77-AEE3-E211-81A0-20CF305B0584.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/B6FEDE40-AEE3-E211-95DB-20CF3027A589.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/C4471CF2-AEE3-E211-8CAC-E0CB4E5536A8.root
root://eospublic.cern.ch/eos/opendata/cms/Run2012B/SingleMu/AOD/22Jan2013-v1/110000/CAD45076-AEE3-E211-8BDB-E0CB4E5536BB.root
```



The screenshot shows a web browser window with the URL `opendata.cern.ch/docs/cms-guide-for-condition-database`. The page title is "CMS Guide to the CMS condition database". There are two tabs: "Documentation" and "Guide", with "Guide" selected. The page content includes:

This page explains the use of global tags and the condition database with the CMS Open Data.

A Global Tag is a coherent collection of records of additional data needed by the reconstruction and analysis software. The Global Tag is defined for each data-taking period, separately for collision and simulated data.

These records are stored in the condition database. Condition data include non-event-related information (Alignment, Calibration, Temperature, etc.) and parameters for the simulation/reconstruction/analysis software. For CMS Open Data, the condition data are provided as sqlite files in the `/cvmfs/cms-opendata-condcdb.cern.ch/` directory, which is accessible through the CMS Open Data VM. Note that when using CMS Open Data docker images, connecting to this area with the command `process.GlobalTag.connect = cms.string(...)` in the job configuration file is not required.

Most [physics objects](#) in the CMS Open Data are already calibrated and ready-to-use, and no additional corrections are needed other than selection and identification criteria, which will be applied in the analysis code. Therefore, simple analyses do not need to access the condition database. Examples of such analyses are [the di-muon spectrum example](#) or [the Higgs analysis example](#).

However, access to the condition database is necessary, for example, for [jet energy corrections](#) and [trigger configuration information](#). Examples of such analyses are for [the PAT object production](#) or [the top quark pair production](#).

Note that when you need to access the condition database, the first time you run the job on the CMS Open Data VM, it will download the condition data from the `/cvmfs` area. It will take time (an example run of a 10 Mbps line took 45 mins), but it will only happen once as the files will be cached on your VM. The job will not produce any output during this time, but you can check the ongoing processes with the command 'top' and you can monitor the progress of reading the condition data to the local cache with the command 'df'.

The Global Tags for condition data are different for different types of data taking. Below, the instructions are given for [proton-proton](#) and

Installation on Debian/Ubuntu

(prepared by V. Olsen - 2017-12-06)

Tested on Debian Stretch

Install Packages

```
$ sudo apt install openafs-client openafs-modules-dkms openafs-krb5 krb5-user krb5-config
```

Configure AFS and Kerberos

1. Use "cern.ch" as default AFS cell

```
$ echo "cern.ch" | sudo tee /etc/openafs/ThisCell
```

2. Set up Kerberos authentication

Add the following lines to file `/etc/krb5.conf`:

```
# settings for CERN.CH realm are taken from file
#  lxplus.cern.ch:/etc/krb5.conf

[libdefaults]
    default_realm = CERN.CH

[realms]
    CERN.CH = {
        default_domain = cern.ch
        kpasswd_server = cerndc.cern.ch
        admin_server = cerndc.cern.ch
        kdc = cerndc.cern.ch
    }

[domain_realm]
    cern.ch = CERN.CH
    .cern.ch = CERN.CH
```

نصب ای.اف.اس

نصب سی.وی.ام

Docs » Welcome to CernVM-FS's documentation! [Edit on GitHub](#)

Welcome to CernVM-FS's documentation!

What is CernVM-FS?

The CernVM-File System (CernVM-FS) provides a scalable, reliable and low- maintenance software distribution service. It was developed to assist High Energy Physics (HEP) collaborations to deploy software on the worldwide- distributed computing infrastructure used to run data processing applications. CernVM-FS is implemented as a POSIX read-only file system in user space (a FUSE module). Files and directories are hosted on standard web servers and mounted in the universal namespace `/cvmfs`. Internally, CernVM-FS uses content-addressable storage and Merkle trees in order to maintain file data and meta-data. CernVM-FS uses outgoing HTTP connections only, thereby it avoids most of the firewall issues of other network file systems. It transfers data and meta-data on demand and verifies data integrity by cryptographic hashes.

By means of aggressive caching and reduction of latency, CernVM-FS focuses specifically on the software use case. Software usually comprises many small files that are frequently opened and read as a whole. Furthermore, the software use case includes frequent look-ups for files in multiple directories when search paths are examined.

CernVM-FS is actively used by small and large HEP collaborations. In many cases, it replaces package managers and shared software areas on cluster file systems as means to distribute the software used to process experiment data.

Contents

- [Release Notes for CernVM-FS 2.8.1](#)

<https://cvmfs.readthedocs.io/en/stable/index.html>

مثال‌ها و نمونه کدها در «گیت‌هاب»

cernopendata / opendata.cern.ch Public

Watch 81 Star 460 Fork 129

Code Issues 142 Pull requests 5 Actions Projects Wiki Security Insights

master 7 branches 0 tags

Go to file Add file Code

ParthS007 facets: update facet counts on every search ✓ 4cd8bd1 19 days ago 2,294 commits

.github	env: set python 3.6 as main supported version	7 months ago
cernopendata	facets: update facet counts on every search	19 days ago
elasticsearch-proxy	elasticsearch: fix reindexing of records	4 years ago
nginx	installation: ssl config and secure headers	8 months ago
scripts	config: add FACET_HIERARCHY variable	7 months ago
sentry	logging: add support for logging to Sentry	4 years ago
tests	search: hide ondemand dataset by default	2 months ago
.dockerignore	tests: addition of pycodestyle	3 years ago
.editorconfig	ci: added github actions workflow	10 months ago
.gitignore	templates: add terms of use -page	4 years ago
.inveniord	docker: nginx static asset exposure fix	4 years ago
AUTHORS.rst	docker: upgrade to rootd 4.12.5	10 months ago
CHANGES.rst	global: package structure and style	4 years ago
CONTRIBUTING.rst	global: package structure and style	4 years ago

About

Source code for the CERN Open Data portal

opendata.cern.ch/

python flask big-data json-schema open-data open-science research-data-management research-data-repository invenio digital-library inveniосоftware open-research-data research-data digital-repository

Readme

GPL-2.0 License

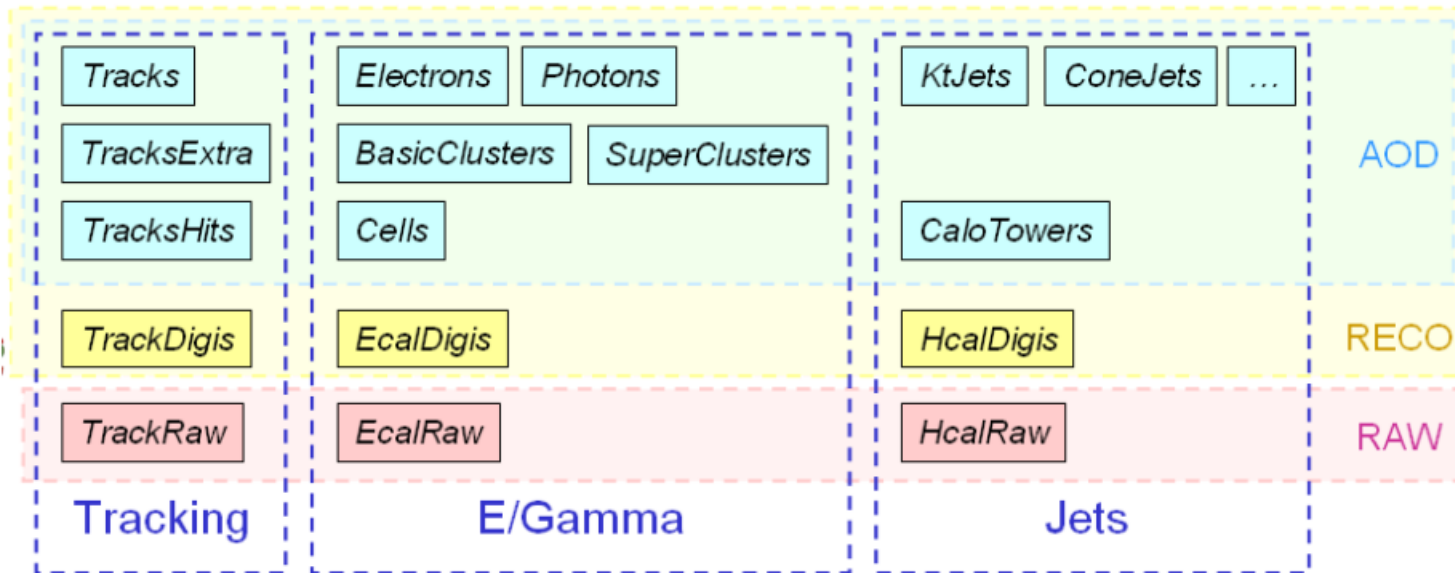
Releases

No releases published

Packages

<https://github.com/cernopendata/opendata.cern.ch>

انواع و محتویات داده‌ها



Physics Analysis Toolkit, or PAT.

Group and user skims: RECO, AOD and PAT-tuples

محتوای داده‌های رویدادهای بازسازی شده

<https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideRecoDataTable>

The screenshot shows a web browser displaying the TWiki page for 'RECO Data Format Table'. The page includes a navigation sidebar on the left with links like 'Log In', 'CMSPublic', 'CMSPublic Web', 'CMSPrivate Web', 'Create New Topic', 'Index', 'Search', 'Changes', 'Notifications', 'Statistics', and 'Preferences'. The main content area features a breadcrumb trail: 'TWiki > CMSPublic Web > SWGuide > SWGuideRecoDataTable (2011-12-15, MantasStankevicius)'. Below the breadcrumb, there are 'Edit', 'Attach', and 'PDF' buttons. The title 'RECO Data Format Table' is followed by 'Complete: 2' and a 'Contents' section with links to 'RECO Data Formats', 'How to use the table', 'Contacts', and 'Information sources'. The main section is titled 'RECO Data Formats' and contains a table with three columns: 'InputTag/Module (Instance name)', 'Containers', and 'Description'. The table is organized into sections: 'Track collections (in RECO and AOD)', 'Track collections (in RECO only)', and 'Local calorimetry reco collections (in RECO only)'. The 'Track collections (in RECO and AOD)' section has one row for 'generalTracks' with container 'reco::TrackCollection'. The 'Track collections (in RECO only)' section has one row for 'generalTracks' with container 'reco::TrackExtraCollection'. The 'Local calorimetry reco collections (in RECO only)' section has five rows for 'hbhereco', 'hfeco', 'horeco', 'zdcereco', and 'castorreco', each with a specific edm container and a description of the collection.

InputTag/Module (Instance name)	Containers	Description
Track collections (in RECO and AOD)		
generalTracks	reco::TrackCollection	Collection of tracks obtained with tracker-standalone reconstruction and officially supported by the Tracker DPG group. Such a collection can contain tracks from different tracking algorithms
Track collections (in RECO only)		
generalTracks	reco::TrackExtraCollection	Track extra for the generalTracks. The trajectory state at the inner and outer most measurements
Local calorimetry reco collections (in RECO only)		
hbhereco	edm::SortedCollection<HBHERecHit>	Joint HCAL barrel+endcap RecHits collection
hfeco	edm::SortedCollection<HFRecHit>	Very Forward calorimeter RecHits collection
horeco	edm::SortedCollection<HOREcHit>	Outer calorimeter RecHits collection
zdcereco	edm::SortedCollection<ZDCRecHit>	Zero-degree calorimeter RecHits collection
castorreco	edm::SortedCollection<CastorRecHit>	Collection of CastorRecHits containing energy deposits for all channels

مشكرم