

Table of Contents

CMS Public Data.....	1
Header page:.....	1
Education page:.....	1
Research page:.....	2
Response to errors.....	7

CMS Public Data

A collection of links to public information on the CMS Open Data program can be found in [CMSPublicDataLinks](#).

From here on, this twiki gives secondary instructions for the use of CMS Open Data, complementary to the primary instructions available on the [CERN Open Data](#) page. Eventually the information will be merged. * the information has been merged to the new portal, so this page is currently obsolete *

Some additional guidance through the [CERN Open Data](#) pages for CMS applications:

Header page:

I want to access information for Educational purposes (general public, school or introductory university level, or preparation of non-experts for Research level)

- choose "start learning"

I want to access information for Research purposes (advanced university and researcher level)

- choose "start analysing"

Education page:

(for additional material also see [this github directory](#))

I want to explore or perform CMS educational exercises for fun or educational purposes

- choose "Learning resources", then
- choose "CMS learning resources", then
- choose your application

If you are a **nonscientist**, a **high school** or **first year university student**, or a **teacher**, you might be particularly interested in the "Physics Masterclasses" exercise.

If you are a **third year university student** or higher, a **scientist**, or a **university teacher**, you might be interested in all the exercises.

I want to play with CMS real data histograms of physics quantities after having obtained a basic notion of what these physics quantities mean

- choose "visualise histograms"
- for instructions click on "Need HELP?"

I want to look at event displays of real events which contributed to the Higgs discovery for fun or educational purposes

- choose "visualise events" (you can exercise the "Need HELP" tutorial on the top right if you like)
- wait for 5 seconds or so, so that the CMS logo hiding the "Open file" button to the left disappears, then click on it,
- choose "Open file(s) from web",
- choose "Education/" and then
- click on [4lepton.ig](#) or [diphoton.ig](#), and wait for the events to load to the right,

- choose one and click on "Load"

alternatively,

- choose "Explore CMS", then
- "CMS derived data sets"
- "more"
- "Higgs candidate events for use in education and outreach". This will display a 4μ , then $4e$, then $2\mu 2e$ event (to see electrons, need to switch them on in menu, and switch off tracks)
- to see Higgs- \rightarrow gamma gamma events, click on "diphoton.ig" below and change settings such that one can see the photons.
- can return to multileptons by clicking on "4lepton.ig"

I want to look at event displays of real J/psi- \rightarrow mumu candidates (and I know what this means)

- as for Higgs (alternatively) above, except, after "more", choose "Dimuon events with invariant mass range 2-5 GeV for public education and outreach"

I want to get a help tutorial how to use the options of the event display

- choose "visualise events", then
- "Need HELP?"

I want to get an introduction to high energy physics and CMS software in order to prepare myself for an activity in the Research section

- choose "Learning resources", then
- choose "CMS learning resources", then
- go to "Computing methods in high energy physics" if you want to learn or refresh the basics of general and CMS computing
- then go to "CMS HEP tutorial" if you want to learn some basics about the analysis of 2011 CMS pp data
- to visualise event displays for the event classes used in the CMS HEP tutorial, follow the 'alternative' instructions for the Higgs event displays above, and choose your favourite sample instead of "more".

I want to explore the sophisticated "Outreach Exercise 2010" featuring an analysis of Z decay to two leptons & ZZ decay to four leptons (needs installation of a virtual machine and software environment as detailed under "Research" below). A video tutorial can be found under * does this actually still work? *

- choose "Explore CMS", then
- choose "CMS open data instructions", then
- "Video tutorial for Outreach Exercise 2010"

I want ... (need to complete the documentation of other entries in the "derived datasets" list, e.g. the analysis of .csv files)

I want to access historic information on 'older' educational datasets released in 2010 and 2011:
CMSPublicData2011.

Research page:

I want to get a general introduction into HEP and CMS software and terminology, with a simplified event format

- go back to the "Education" section (see above) and follow the corresponding exercises

I want to learn about the terms under which I can access and use the CMS Open Data, and publish results obtained from them

- choose "Explore CMS", then (in text description)
- "About CMS", then (at bottom of page)
- "Data preservation and open access policy" (if you are a CMS member see also "papers by CMS members using public data")

I want to get inspiration for some potential physics topics

- a link with examples of potentially interesting physics topics and their relation to CMS open data might be added here soon.

I want to learn about the nature of the CMS physics objects and the corresponding variables and terminology

- choose "Explore CMS", then (in text description)
- "About CMS" and
- "About CMS physics objects"

I want to find out whether I should go for 2010 or 2011 data (both are pp data at 7 TeV)

- the 2010 data have been released first, have fewer, smaller data sets, better low pt tracking, low trigger thresholds, low pileup, and more/simpler analysis/validation examples, but no MC. If you do not need MC or maximal statistics, you might want to try 2010 data first (simpler).
- the 2011 data have more statistics, more diverse data sets, many associated MC sets, and a slightly more advanced VM environment. If you are immediately interested in maximal statistics and/or MC acceptance corrections you should go for 2011 data (more sophisticated).
- information on the respective luminosities and pileup rates vs. time can be found here.
- there is nothing wrong with trying both (on separate VM's).

I want to install the CMS virtual machine (separately for 2010 and 2011 data) which is needed for CMS Research level data analysis.

- choose "Install your virtual machine"
- choose "CMS virtual machines"

or

- choose "Explore CMS"
- choose "VMs"

I want to install the CMS software environment on the virtual machine (separately for 2010 and 2011 data) which is needed for access to and analysis of CMS Research level data.

- choose "Start analysing the data"
- choose "CMS Getting Started"
- choose "2011" (default) or "2010"

or

- choose "Explore CMS"

- choose "Getting started!"
- choose "2011" (default) or "2010"

The 2010 (SL5) virtual machine will only work on 2010 data with CMSSW 4-2-8. The 2011 (SL6) virtual machine will only work on 2011 data and MC with CMSSW 5-3-32.

I want to produce some example physics distributions (inclusive dimuon spectrum analysis example directly from AOD or two lepton/four lepton analysis example with intermediate ntuples)

- Install the CMS software as in the previous item (for 2010 or 2011)
- follow options A (dimuon spectrum, simpler) or B (two/four lepton example, more sophisticated). For 2011 option A, take the SingleMu or DoubleMu dataset index files from <http://opendata.cern.ch/record/32> or <http://opendata.cern.ch/record/17>

I want to find out which 2010 data sets exist, and how to get a feel for their content

- choose "Explore CMS"
- in "CMS primary datasets", choose "2010"
- choose a data set, and read comments

and/or, to view some corresponding event displays

- go back to "Education"
- choose "Explore CMS", then
- in the box "CMS derived data sets" click on "2010"
- choose "Event display file derived from" ... the name of the CMS primary data set you want (ZeroBias is known to be essentially empty)

alternative way to view these event displays

- in Education section, choose "visualise events"
- choose "Open File", "Open Files from Web", "2010"
- choose your data set

I want to find out which 2010 data set and/or analysis/validation example is most useful for my purpose

- to learn how to do a muon analysis, follow **I want to produce some first physics distributions**, options A (recommended) or B, or try one of the relevant **I want to (re)validate the 2010 data sets** examples
- to learn how to do an electron analysis, follow **I want to produce some first physics distributions**, option B, or try one of the relevant **I want to (re)validate the 2010 data sets** examples
- to learn how to do a minimum bias track analysis, try the MinimumBias example on **I want to (re)validate the 2010 data sets**
- more will follow

I want to find out which 2011 data and MC sets exist, and how to get a feel for their content

- choose "Explore CMS"
- in "CMS primary datasets", Or "CMS simulated data sets", choose "2011"
- choose a data set, and read comments

and/or, to view some corresponding event displays (data only for the time being)

- go back to "Education"

- choose "Explore CMS", then
- in the box "CMS derived data sets" click on "2011"
- choose "Event display file derived from"... the name of the CMS primary data set you want (ZeroBias is known to be essentially empty)

alternative way to view these event displays (data only)

- in Education section, choose "visualise events"
- choose "Open File", "Open Files from Web", "2011"
- choose your data set

how to interpret the MC set names?

- check description [↗](#)

I want to find out which 2011 data set and/or analysis/validation example is most useful for my purpose

- dedicated examples for 2011 beyond those available in "Getting Started" are in preparation [↗](#) to be added to the portal, including jet and top cross sections. Alternatively, start from a 2010 example and adjust to run on 2011 data.

I want to find out how to use the trigger and trigger prescale information in the data set I am interested in (still very basic, to be improved)

- choose "Explore CMS"
- choose "CMS Trigger Information" (currently for 2011 only)

I want to find out how to access the luminosity information for the data set I am interested in and how to select "good data" only

- check the luminosity record [↗](#)
- check the JSON description [↗](#)

I want to find out whether I need condition data base information, and if so, how to access it (still very basic, to be improved)

- condition data are needed only on sophisticated examples using e.g. jet energy corrections (many of the simpler analysis/validation examples documented here do not)
- using condition data significantly slows down data access, so use them only if really needed. If so:
- choose "Explore CMS"
- choose "CMS Condition Data" for basic information (in contrast to what is stated there, they are NOT needed to run CERNVM in general, only to perform sophisticated tasks)

I want to find more CMS software and data format documentation from public sources (strongly recommended for serious analysis, but hard to navigate!)

- check the CMS public web or the CMS Fermilab website or Google

I want to use the external public analysis example from the MIT jet analysis group papers

- Jet Substructure Studies with CMS Open Data, A. Tripathy et al., Apr 19, 2017, MIT-CTP-4890, arXiv:1704.05842 [↗](#)
- Exposing the QCD Splitting Function with CMS Open Data, A. Larkoski et al., Apr 17, 2017, MIT-CTP-4891, arXiv:1704.05066 [↗](#)

I want to (re)validate the 2010 data sets within my setup

for the MinimumBias, Commissioning, Mu or MuMonitor data sets:

- Explore CMS
- on "Validation Utilities", choose "2010"
- choose and execute the corresponding Validation code

for the Multijet data set: (see also here [?](#))

- Explore CMS
- on "CMS Tools", choose "2010", then
- choose and execute "Razor filter and analyzer for SUSY searches"

for the Electron (or Mu) data set:

- Explore CMS
- on "CMS Tools", choose "2010"
- choose and execute "Software to preprocess the CMS 2010 Muon and Electron datasets for the two-lepton/four-lepton analysis example of CMS open data", then
- choose "Two-lepton/four-lepton analysis example of CMS 2010 open data" and compare PAT-tuples from the previous to those linked therein, or execute it on your new PAT tuples

for the ZeroBias data set:

- not useful, no validation needed

for the Jet, MuOnia, BTau, Photon, JetMETTaumonitor, METFwd data sets:

- validation not yet available (partially in preparation)

I want to backup my code, or import some external code

- we recommend to use scp from and to your host from within the VM

I want to find the luminosity of my data set, possibly constrained by using specific triggers

- please check CMS luminosity information [?](#), i.e.
- decide on the triggers you want to use
- find out the runs/lumi sections in which these triggers were active and whether they were prescaled
- overlap with the available Open Data samples (run range) and with the JSON data quality selection
- if not prescaled, sum up the luminosity for the surviving runs/lumi sections
- if prescaled, life is more complicated ...

I want to find the effective luminosity of my MC set

- to be documented.
- generically: divide MC cross section (next item) times matching efficiency times filter efficiency by number of events.

I want to find the generator cross section of a particular MC set

- to be documented ...
- on some MC sets, the following might work (reliability of information not guaranteed): open the

ROOT file, create TBrowser, navigate to: Runs -> GenRunInfoProduct_generator__SIM. -> GenRunInfoProduct_generator__SIM.obj -> InternalXSec -> value_

I want ... (a few other already existing things are not yet documented)

I want other information than the one documented here and on <http://opendata.cern.ch>

- kindly contact opendatasupport@cernNOSPAMPLEASE.ch

Response to errors

I have trouble installing Virtualbox

- read the FAQs. If it still fails, please contact your local system administrator

When/after installing CERNVM, I get a message that my VM uses too much memory

- reduce the memory allocated to the virtualbox

When reading AOD data, I get write access warnings from eospublic on every file

- this was a temporary 'feature' of a change in the eos software which has meanwhile been fixed. It does not affect the results.

When reading AOD data, I get fatal access error messages from eospublic on specific files

- According to the eos-admin team, disk access problems to eospublic may occur occasionally and are automatically corrected within a few hours. If the access problem to a particular file lasts longer than about a day send a mail to eos-admins@cernNOSPAMPLEASE.ch, providing the file name and a log of the error message.

While running on AOD data, my job gets "killed" by the VM without any further explanation

- You might have exceeded the VMs available memory (use the VMs monitoring tools to check whether memory is marginal). Try **one** of the following:
 - do not run anything else using a lot of memory (e.g. a web browser) on the VM in parallel
 - reduce memory usage of the job (and/or check for potential memory leak)
 - increase the memory allocated to the VM

Any other problem you cannot solve yourself or with the help of your local administrator(s), not related to your local setup

- kindly contact opendatasupport@cernNOSPAMPLEASE.ch

-- AchimGeiser - 2017-04-25

This topic: CMSPublic > CMSPublicData

Topic revision: r11 - 2018-01-15 - AchimGeiser



Copyright &© 2008-2020 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback