

CERN Open Data Portal

Tibor Šimko

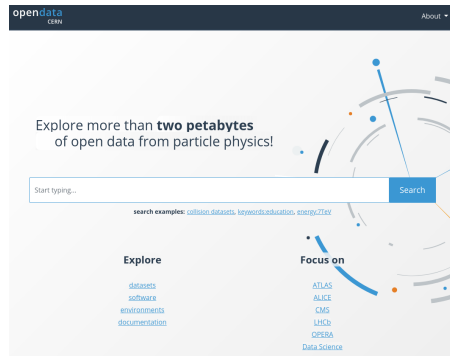
CERN

3rd DPHEP Collaboration Workshop, 21–23 June 2021

<https://indico.cern.ch/event/1043155>

CERN Open Data portal

- ▶ launched in November 2014
- ▶ rich content
 - ▶ derived datasets for education
 - ▶ collision and simulated datasets for research
 - ▶ configuration files and documentation
 - ▶ virtual machines and container images
 - ▶ software tools and analysis examples
- ▶ total size in June 2021
 - ▶ over 7'600 bibliographic records
 - ▶ over 900'000 files
 - ▶ over 2.4 petabytes

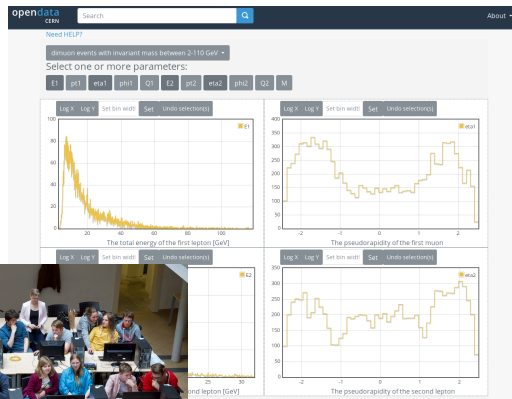
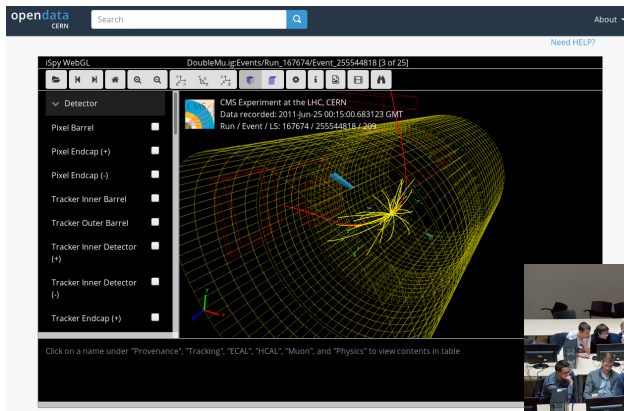


<https://opendata.cern.ch>

Developed by CERN in close collaboration with Experiments



Education-oriented use cases



Interactive event display and histogramming for derived datasets

Latest education-level content news

open.data CERN

ATLAS simulated samples collection for jet reconstruction training, as part of the 2020 Open Data release

ATLAS Collaboration

Cite as: ATLAS Collaboration (2020), ATLAS simulated samples collection for jet reconstruction training, as part of the 2020 Open Data release. CERN Open Data Portal. DOI:10.14602/OPENDATA.ATLAS.18000000

Home About Downloads

Description

A set of dedicated Monte-Carlo simulated samples corresponding to Standard Model hadronic physics processes, by the ATLAS Collaboration for educational purposes related to Jet Clustering techniques and reconstruction algorithms.

Dataset characteristics

1629600 events, 19 files, 22.3 GB in total.

How were these data selected?

Dataset designed to teach jet reconstruction. See (1) files of 102.9k events each (for a total 1.629 million of events). Each file is 343 bins with 300 events each. The set corresponds to simulated (Monte Carlo) data only. For more details of the samples, please refer to the reference below.

How can you use these data?

This dataset is provided by the ATLAS Collaboration **only for educational purposes and is not suited for scientific publications**. More documentation is available at:

The ATLAS Open Data documentation for this and other datasets

The ATLAS Open Data site

Files

Filename	Size
mc_jets-2020-p4e01.root	2.1 GB
mc_jets-2020-p4e02.root	2.1 GB
mc_jets-2020-p4e03.root	2.1 GB
mc_jets-2020-p4e04.root	2.1 GB
mc_jets-2020-p4e05.root	2.1 GB

First Previous 1 2 Next Last

- ▶ **May 2021** ATLAS releases jet reconstruction training samples

open.data CERN

Examples of basic analysis techniques for neutral meson and photon data from the PHENIX detector

@ David Goussier @ Prasanna Mooney PHENIX collaboration

Cite as: David Goussier, Prasanna Mooney, PHENIX collaboration (2021), Examples of basic analysis techniques for neutral meson and photon data from the PHENIX detector. CERN Open Data Portal. DOI:10.14602/OPENDATA.PHENIX.10000007

Home About Downloads

Description

This record contains datasets from the Electromagnetic Calorimeter (EMCAL) of the PHENIX detector. It aims to present a few basic techniques of identifying π^0 s and photons using that device. The data is provided in the form of ROOT files containing small samples (hundreds) of reconstructed $p+K_u$ data taken with the PHENIX detector in 2015 at $\sqrt{s_{NN}}=200$ GeV. The Gamma-Gamma Nucleus (ggNucleus) and Single Quark Nucleus (sqN) are provided. Both minimum bias and EMCAL-triggered data are included. These samples were created for educational and training purposes only and are not fit for use for deriving any real-physicist results. An example code working with the data is attached as ROOT macros. A detailed guide is also provided.

Dataset characteristics

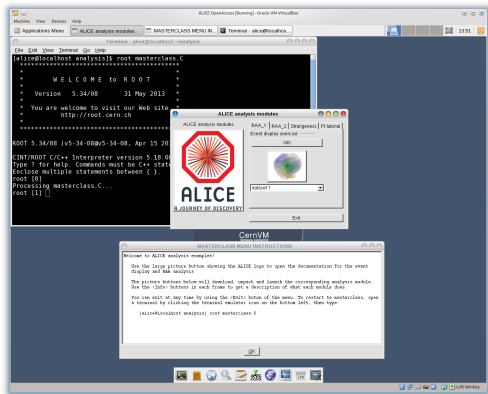
1255308 events, 9 files, 73.7 MB in total.

Dataset semantics

Variable	Type	Description
event	float	Event centrality
vertexZ	float	z-vertex of the event
pt	float	Transverse momentum of cluster pair (in candidates) for ggNucleus; Transverse momentum of the cluster (y-candidates) for sqN
costheta	float	cos of the opening angle between the two clusters for ggNucleus; Polar angle of the cluster (y-candidates) for sqN
phi	float	Azimuthal angle of the direction of the pair's momentum (assumed 90°) for ggNucleus; Azimuthal angle of the cluster (y-candidates) for sqN
mass	float	Invariant mass calculated from the two clusters (energy and position)
asym	float	Energy asymmetry $(E1-E2)/(E1+E2)$ of the two clusters
sect	float	EMCAL sector where the first cluster is
Ecorr1	float	"Corrected" energy of the first cluster
phi1	float	True phi angle of the first cluster

- ▶ **June 2021** First open data release from the PHENIX experiment

Research-oriented use cases



Run CernVM Virtual Machines

open data
CERN

Search

About

Higgs-to-four-lepton analysis example using 2011-2012 data

Jomhari, Nur Zulaiha; Geiser, Achim; Bin Anuar, Afiq Alzuddin

Cite as: Jomhari, Nur Zulaiha; Geiser, Achim; Bin Anuar, Afiq Alzuddin; (2017). Higgs-to-four-lepton analysis example using 2011-2012 data. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS_JKBS.RR4Z

Software Analysis CMS Accelerator CERN.LHC

Description

This research level example is a strongly simplified reimplement of parts of the original CMS Higgs to four lepton analysis published in [Phys.Lett. B716 \(2012\) 30-61, arXiv:1207.7235](#).

The published reference plot which is being approximated in this example is https://inspirehep.net/record/1124338/files/H41_mass_3.png. Other Higgs final states (e.g. Higgs to two photons), which were also part of the same CMS paper and strongly contributed to the Higgs boson discovery, are not covered by this example.

The example consists of different levels of complexity. The highest level of this example addresses users who feel they have at least some minimal understanding of the content of this paper and of the meaning of this reference plot, which can be reached via (separate) educational exerc with the linux op

Use with

The example uses publication due to but not identical to in many later CM

/DoubleElectron/
/DoubleMu/Run2

Events / 3 GeV

m_4 [GeV]

Events / 3 GeV

m_4 [GeV]

re original again close to, ly as they are,

Run simplified research-level analysis examples

... enables independent theoretical research

Welcome to INSPIRE, the High Energy Physics information system. Please direct questions, comments or concerns to feedback@inspirehep.net

HEP :: HEPNames :: Institutions :: Conferences :: Jobs :: Experiments :: Journals :: Help

reference:10.7483/OPENDATA.CMS Brief format Search New Search Advanced Search

Sort by: earliest date Display results: 25 results

No exact match found for 10.7483/OPENDATA.CMS, using 10.7483/OPENDATA.CMS instead...

HEP 42 records found 1 - 25 jump to record: 1 Search took 0.13 seconds.

1. Fast and Accurate Simulation of Particle Detectors Using Generative Adversarial Networks

⁽⁴²⁾ Pasquale Musella (ETH, Zurich (main)), Francesco Pandolfi (INFN, Rome), May 2, 2018, 8 pp.
Published in *Comput. Softw. Big Sci.* **2** (2018) no.1, 8
DOI: [10.1007/s41781-018-0015-y](https://doi.org/10.1007/s41781-018-0015-y)
e-Print: [arXiv:1805.09850](https://arxiv.org/abs/1805.09850) [hep-ex] | PDF

References | BibTeX | LaTeXJUS | LaTeX(EU) | HarvMatic | EndNote
ADS Abstract Service

Detailed record - Cited by 42 records

2. Exposing the QCD Splitting Function with CMS Open Data

⁽³³⁾ Andrew Larkoski (Reed Coll.), Simone Marzani (SUNY, Buffalo), Jesse Thaler, Aashish Tripathi, Wei Xue (MIT, Cambridge, CTP), Apr 17, 2017, 7 pp.
Published in *Phys.Rev.Lett.* **119** (2017) no.13, 132003

MIT-CTP-4891
DOI: [10.1103/PhysRevLett.119.132003](https://doi.org/10.1103/PhysRevLett.119.132003)
e-Print: [arXiv:1704.05066](https://arxiv.org/abs/1704.05066) [hep-ph] | PDF

References | BibTeX | LaTeXJUS | LaTeX(EU) | HarvMatic | EndNote
ADS Abstract Service

Detailed record - Cited by 39 records

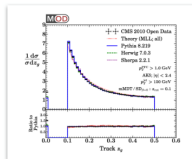
3. Reinterpretation of LHC Results for New Physics: Status and Recommendations after Run 2

⁽³¹⁾ LHC Reinterpretation Forum Collaboration (Waleed Abdallah (Harish-Chandra Res. Inst. & Cairo U) et al.), Mar 19, 2020, 58 pp.
Published in *SciPost Phys.* **9** (2020) no.2, 022

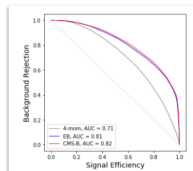
CERN-LPCC-2020-001, FERMILAB-FN-1098-CMS-T, Imperial/HEP/2020/RIF/01
DOI: [10.21468/SciPostPhys.9.2.022](https://doi.org/10.21468/SciPostPhys.9.2.022)
e-Print: [arXiv:2003.07988](https://arxiv.org/abs/2003.07988) [hep-ph] | PDF

References | BibTeX | LaTeXJUS | LaTeX(EU) | HarvMatic | EndNote
CERN Document Server | ADS Abstract Service | OSTI.gov Server | Link to Fermilab Library Server (fulltext available) | Link to Fulltext from Publisher | Link to Fulltext

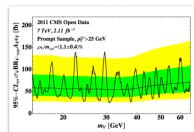
Detailed record - Cited by 21 records



arXiv:1704.05066



arXiv:1807.11916



arXiv:1902.04222

Searches, QCD jet studies, Machine Learning...

arXiv:1708.09429v2

A measurement of the z_b distribution in pp collisions, using CMS open data, was recently reported [33, 34]. In PbPb collisions, this measurement reflects how the two color-charged par-

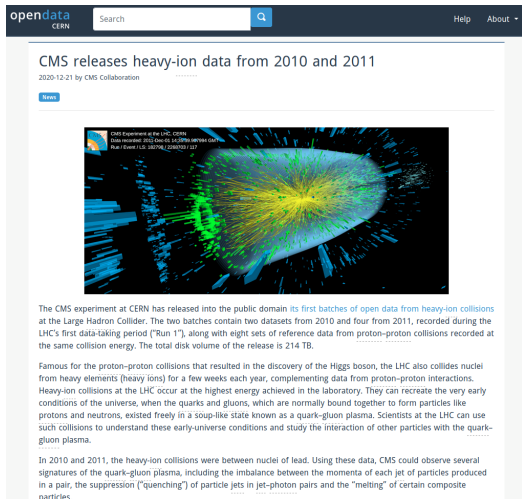
[33] A. Larkoski et al., “Exposing the QCD splitting function with CMS Open Data”, *Phys. Rev. Lett.* **119** (2017) 132003, [doi:10.1103/PhysRevLett.119.132003](https://doi.org/10.1103/PhysRevLett.119.132003), [arXiv:1704.05066](https://arxiv.org/abs/1704.05066).

[34] A. Tripathi et al., “Jet Substructure Studies with CMS Open Data”, *Phys. Rev. D* **96** (2017) 074003, [doi:10.1103/PhysRevD.96.074003](https://doi.org/10.1103/PhysRevD.96.074003), [arXiv:1704.05842](https://arxiv.org/abs/1704.05842).

... that the CMS collaboration cites!

Over forty papers citing CMS open data

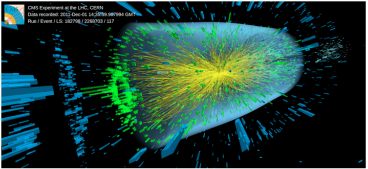
Latest research-level content news



The screenshot shows the OpenData CERN website interface. At the top, there is a search bar and navigation links for 'Help' and 'About'. The main content area features a news article with the title 'CMS releases heavy-ion data from 2010 and 2011' and a sub-headline '2020-12-21 by CMS Collaboration'. Below the text is a 'News' tag and a large image of a particle collision event. The image shows a central bright yellow and green core with numerous blue and green tracks radiating outwards, representing the tracks of particles produced in a heavy-ion collision. The image is overlaid with a semi-transparent blue and green shape, possibly representing the collision region or a specific particle track.

CMS releases heavy-ion data from 2010 and 2011
2020-12-21 by CMS Collaboration

News



The CMS experiment at CERN has released into the public domain its first batches of open data from heavy-ion collisions at the Large Hadron Collider. The two batches contain two datasets from 2010 and four from 2011, recorded during the LHC's first data-taking period ("Run 1"), along with eight sets of reference data from proton-proton collisions recorded at the same collision energy. The total disk volume of the release is 214 TB.

Famous for the proton-proton collisions that resulted in the discovery of the Higgs boson, the LHC also collides nuclei from heavy elements (heavy ions) for a few weeks each year, complementing data from proton-proton interactions. Heavy-ion collisions at the LHC occur at the highest energy achieved in the laboratory. They can recreate the very early conditions of the universe, when the quarks and gluons, which are normally bound together to form particles like protons and neutrons, existed freely in a soup-like state known as a quark-gluon plasma. Scientists at the LHC can use such collisions to understand these early-universe conditions and study the interaction of other particles with the quark-gluon plasma.

In 2010 and 2011, the heavy-ion collisions were between nuclei of lead. Using these data, CMS could observe several signatures of the quark-gluon plasma, including the imbalance between the momenta of each jet of particles produced in a pair, the suppression ("quenching") of particle jets in jet-photon pairs and the "melting" of certain composite particles.

- ▶ **December 2020** CMS released first 2010–11 heavy-ion data samples and reference proton-proton datasets (214 TB).

<https://opendata.cern.ch/docs/cms-releases-heavy-ion-data>

- ▶ **October 2020** CMS run a CMS Open Data Workshop for Theorists at the LPC.

<https://indico.cern.ch/event/882586/>

- ▶ **August 2020** CMS released fifth batch of new open data. All proton-proton collision data recorded in 2010–11 and half of 2012 are available throughout the portal.

<https://opendata.cern.ch/docs/cms-completes-2010-2011-pp-data>

FAIR guiding principles for scientific data management

scientific **data**

[Explore Content](#) ▾ [Journal Information](#) ▾ [Publish With Us](#) ▾

[nature](#) > [scientific data](#) > [comment](#) > [article](#)

[Open Access](#) | [Published: 15 March 2016](#)

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons [✉](#)

Scientific Data **3**, Article number: 160018 (2016) | [Cite this article](#)

180k [Accesses](#) | 2306 [Citations](#) | 1797 [Altmetric](#) | [Metrics](#)

i An Addendum to this article was published on 19 March 2019

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

- ▶ Findable
- ▶ Accessible
- ▶ Interoperable
- ▶ Reusable

Findable: persistent identifiers & machine readability

The screenshot shows the OpenData CERN interface. At the top, there is a search bar and an 'About' link. The main content area displays the title 'Mu primary dataset in AOD format from RunB of 2010 (/Mu/Run2010B-Apr21ReReco-v1/AOD)' and the source 'CMS collaboration'. Below this, there are tabs for 'Dataset', 'Collisions', 'CMS', 'Collision energy /TeV', 'Accelerator/CERN/LHC', and 'Parent Dataset: /Mu/Run2010B-v1/RAW'. The 'Description' section states 'Mu primary dataset in AOD format from RunB of 2010'. The 'Notes' section explains that the dataset contains all runs from 2010 RunB and provides a link to a CMS list of validated runs. The 'Related Datasets' section shows a link to '/Mu/Run2010B-v1/RAW'. The 'Characteristics' section lists 'Dataset: 32376291 events 2979 files 3.2 TB in total'. The 'System Details' section includes a global tag 'FT_R_42_V10A:All' and a recommended release for analysis: 'CMS5W_A_2_1_patch1'.

```
{
  "created": "2020-12-21T10:16:53.741747+00:00",
  "id": 14,
  "metadata": {
    "schema": "http://opendata.cern.ch/schema/records/record-v1.0.0.json",
    "abstract": {
      "description": "14:Mu primary dataset in AOD format from RunB of 2010: /p <br/> This dataset contains all runs from 2010 RunB. The list of validated runs, which must be applied to all analyses, can be found in: /p",
      "links": [
        {
          "recid": "1009"
        }
      ]
    }
  },
  "accelerator": "CERN-LHC",
  "collaboration": {
    "name": "CMS collaboration",
    "recid": "459"
  },
  "collections": [
    "CMS-Primary-Datasets"
  ],
  "collision_information": {
    "energy": "13",
    "type": "pp"
  },
  "control_number": "14",
  "date_created": {
    "2010"
  },
  "date_published": "2014",
  "date_reprocessed": "2011",
  "distribution": {
    "formats": [
      "root",
      "aod"
    ]
  },
  "number_events": 32376291,
  "number_files": 2979,
  "size": 326262517610
},
{
  "doi": "10.7483/OPENDATA.CMS.B8MR.C4A2",
  "experiment": "CMS",
  "files": [],
  "index_files": [
    {
      "checksum": "adler32:85e137b0",
      "filename": "CMS_Run2010B_Mu_AOD_Apr21ReReco-v1_0000_file_index.json",
      "size": 180230,
      "url_http": "http://opendata.cern.ch/record/14/files/CMS_Run2010B_Mu_AOD_Apr21ReReco-v1_0000_file_index.json",
      "url_root": "root://eospublic.cern.ch/eosopendata/cms/Run2010B/MuAOD/Apr21ReReco-v1/1file-Indexes/CMS_Run2010B_Mu_AOD_Apr21ReReco-v1_0000_file_index.json"
    },
    {
      "checksum": "adler32:1869e093",
      "filename": "CMS_Run2010B_Mu_AOD_Apr21ReReco-v1_0000_file_index.txt",
      "size": 85184,
      "url_http": "http://opendata.cern.ch/record/14/files/CMS_Run2010B_Mu_AOD_Apr21ReReco-v1_0000_file_index.txt"
    }
  ]
}
```

<https://opendata.cern.ch/record/14> \equiv <http://doi.org/10.7483/OPENDATA.CMS.B8MR.C4A2>

Each dataset is identified by a “record ID” and optionally minted with a DOI

Findable: rich context description & machine readability

How were these data selected?

There are four categories of triggers in the Mu dataset (with significant overlaps):

-70% inclusive single muon triggers with varying trigger pt threshold 3,5,7,9,11,13,15,17,19,21 GeV plus a few with loosened quality cuts.

-20% isolated single muon triggers with varying trigger pt threshold 9,11,13,15,17 GeV.

-10% inclusive dimuon triggers with varying trigger pt threshold 3,5 GeV plus one Z-muon trigger with loosened quality cuts.

-20% combinations of muon triggers with various pt thresholds 3,5,7,8,9,11 GeV with some EM/e/gamma or hadronic/jet energy deposit with thresholds 6-100 GeV.

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 Virtual Machine. See the instructions for setting up the Virtual Machine and getting started in:

[How to install the CMS Virtual Machine](#)

[Getting started with CMS open data](#)

```
  }
  }
  "run_period": {
    "run2010B"
  }
  "system_details": {
    "global_tag": "FT_6_42_V10A:ATL",
    "release": "CMS_4_2_8"
  },
  "title": "Mu/Run2010B-Apr11Reco-v1/AOD",
  "title_additional": "Mu primary dataset in AOD format from RunB of 2010 (/Mu/Run2010B-Apr11Reco-v1/AOD)",
  "type": {
    "primary": "Dataset",
    "secondary": [
      "Collision"
    ]
  },
  "usage": {
    "description": "You can access these data through the CMS Virtual Machine. See the instructions for setting up the Virtual Machine and getting started in",
    "links": [
      {
        "description": "How to install the CMS Virtual Machine",
        "url": "/docs/cms-virtual-machine-2010"
      },
      {
        "description": "Getting started with CMS open data",
        "url": "/docs/cms-getting-started-2010"
      }
    ]
  },
  "validation": {
    "description": "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:"
  },
  "links": [
    {
      "description": "CMS data quality monitoring: Systems and experiences",
      "url": "http://opendata.cern.ch/record/143899/files/CR2013_410.pdf"
    },
    {
      "description": "The CMS Data Quality Monitoring software experience and future improvements",
      "url": "http://cds.cern.ch/record/143899/files/CR2013_410.pdf"
    },
    {
      "description": "The CMS data quality monitoring software: experience and future prospects",
      "url": "http://opendata.cern.ch/record/143899/files/CR2013_410.pdf"
    }
  ]
},
"updated": "2020-12-21T10:56:54.170861+00:00"
}
```

<https://opendata.cern.ch/record/14> \equiv <http://doi.org/10.7483/OPENDATA.CMS.B8MR.C4A2>

Rich curated context information on data selection, validation, and use

Accessible: downloading content

The screenshot shows the Open Data Client interface. At the top, there is a 'List of files' window with a table of files and their sizes. Below this, there is a 'File Indexes' section with a table of file names and sizes. The interface also includes navigation buttons and a disclaimer.

File Name	Size
00E16FBB-9071-E011-83D3-003048673F12.root	583.0 MB
0248915F-EE71-E011-8894-0025902009E8.root	677.0 MB
0268F635-B671-E011-9090-002481E14E00.root	822.5 MB
0278F65A-9A71-E011-A5C0-0025902008A8.root	630.2 MB
02FF3E00-C171-E011-84C7-002590200ADC.root	766.5 MB

Filename	Size
CMS_Run20108_8TeV_ADD_Apr21ReReco-v1_0000_8TeV_index.txt	75.1 kB
CMS_Run20108_8TeV_ADD_Apr21ReReco-v1_0001_8TeV_index.txt	42.0 kB
CMS_Run20108_8TeV_ADD_Apr21ReReco-v1_0002_8TeV_index.txt	62.7 kB
CMS_Run20108_8TeV_ADD_Apr21ReReco-v1_0003_8TeV_index.txt	54.3 kB
CMS_Run20108_8TeV_ADD_Apr21ReReco-v1_0004_8TeV_index.txt	55.7 kB



The cernpendata-client is a command-line tool to interact with the CERN Open Data portal.

Navigation

1. Installation
2. Usage
3. CLI API
4. Changes
5. Contributing
6. License
7. Authors

cernpendata@GitHub
cernpendata@Twitter
opendata-forum.cern.ch
opendata.cern.ch

Quick search

cernpendata-client

version 2.7.1.0, 1.7.1, 0.8.1.0, 0.7.0, 0.6.0.0, 0.5.0.0, 0.4.0.0, 0.3.0.0, 0.2.0.0, 0.1.0.0, 0.0.1.0, 0.0.0.0

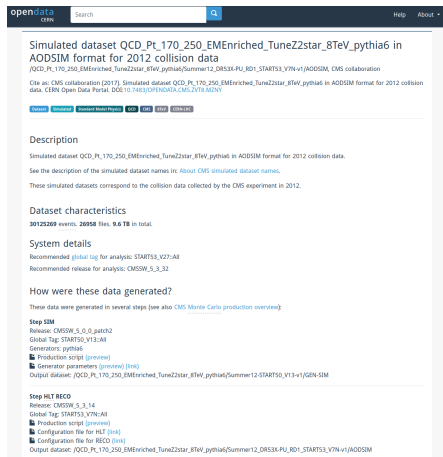
cernpendata-client is a command-line tool to facilitate downloading files from the CERN Open Data portal. The tool enables to query datasets hosted on the CERN Open Data portal and to download and verify the individual data set files.

1. Installation
 - 1.1. PyPI
 - 1.2. Docker
2. Usage
 - 2.1. General help
 - 2.2. Selecting records
 - 2.3. Getting metadata
 - 2.4. Listing available data files
 - 2.5. Downloading data files
 - 2.6. Verifying files
 - 2.7. Listing directories
 - 2.8. More information
3. CLI API
 - 3.1. cernpendata-client
 - 3.1.1. download-files
 - 3.1.2. get-file-locations
 - 3.1.3. get-metadata
 - 3.1.4. list-directory
 - 3.1.5. verify-files
 - 3.1.6. version
4. Changes
 - 4.1. Version 0.2.0 (2020-11-19)
 - 4.2. Version 0.1.0 (2020-09-24)
 - 4.3. Version 0.0.1 (2020-09-09)
5. Contributing
 - 5.1. Issues
 - 5.2. Pull requests
6. License
7. Authors

Browsing and downloading files manually

► September 2020 Automated client

Accessible: working with large datasets



The screenshot shows the OpenData CERN website interface. At the top, there is a search bar with the text "opendata CERN" and a search icon. Below the search bar, the page title is "Simulated dataset QCD_Pt_170_250_EMEnriched_TuneZ2star_8TeV_pythia6 in AODSIM format for 2012 collision data". The page content includes a description of the dataset, its characteristics (30125209 events, 26958 files, 9.6 TB in total), system details (recommended global tag for analysis: START53_V27:All), and how the data was generated (Step SIM and Step HLT RECO).

A dataset of 9.6 TB size, 26'958 files

```
/tmp $ cernopendata-client download-files --help
Usage: cernopendata-client download-files [OPTIONS]
```

Download data files belonging to a record.

Select a CERN Open Data bibliographic record by a record ID, a DOI, or a title and download data files belonging to this record.

Examples:

```
$ cernopendata-client download-files --recid 5500
$ cernopendata-client download-files --recid 5500 --filter-name BuildFile.xml
$ cernopendata-client download-files --recid 5500 --filter-regex py$
$ cernopendata-client download-files --recid 5500 --filter-range 1-4
$ cernopendata-client download-files --recid 5500 --filter-range 1-2,5-7
$ cernopendata-client download-files --recid 5500 --filter-regex py --filter-range 1-2
```

```
/tmp lm 4s $ cernopendata-client get-file-locations --recid 8884 | head -3
http://opendata.cern.ch/eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/QCD_Pt_v1/00000/0052F9A9-5EB2-E311-81A9-0026189437ED.root
http://opendata.cern.ch/eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/QCD_Pt_v1/00000/00BF7DF8-69B2-E311-89E6-003048678E8A.root
http://opendata.cern.ch/eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/QCD_Pt_v1/00000/020194EC-CAB2-E311-8B3A-003048678B94.root
```

Downloading file locations

```
/tmp $ cernopendata-client download-files --recid 8884 --filter-range 1
==> Downloading file 1 of 1
-> File: ./8884/0052F9A9-5EB2-E311-81A9-0026189437ED.root
[ ] -> Progress: 38869/150643 KiB (25%)
```

Downloading and verifying files

Interoperable: data semantics

opendata CERN Search Help About

Samples with full event information including tracker hits for tracking, ML, and top quark tagging studies

Usal, Emanuelle; Andrews, Michael; Burkle, Bjorn; Gleyzer, Sergei; Narain, Meenakshi

Cite as: Usal, Emanuelle; Andrews, Michael; Burkle, Bjorn; Gleyzer, Sergei; Narain, Meenakshi (2019). Samples with full event information including tracker hits for tracking, ML, and top quark tagging studies. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.CHC3.SKPG

Dataset: Tracker hit-enriched 300 to 600 bin of QCD, Pt-15to3000, TuneZ2star, Flat, 8TeV, pythia6
Dataset: Tracker hit-enriched 400 to 600 bin of QCD, Pt-15to3000, TuneZ2star, Flat, 8TeV, pythia6
Dataset: Tracker hit-enriched 600 to 3000 bin of QCD, Pt-15to3000, TuneZ2star, Flat, 8TeV, pythia6
Dataset: Tracker hit-enriched TJets, HadronicMGDecays, 8TeV, madgraph

Description

Samples in this record are in a custom root ntuple format and contain the position of the hits and information from the generator-level objects associated to the tracker hits. The samples can be used to study top quark identification algorithms that use low-level detector information such as tracker hits. Machine learning algorithms are suitable for this classification task.

They have been produced from datasets, which consists of events extracted from simulated proton-proton collision events at a center-of-mass energy of 8 TeV generated with Pythia 6 (QCD) or MadGraph2.6 and Pythia6 (top-antitop pair sample). The particles emerging from the collisions traverse through a simulation of the CMS detector.

The parent datasets of these samples contain light jets (QCD) in various energy ranges or all-hadronic high transverse momentum decays of top quarks and consist of hits from the tracking detector, reconstructed tracks, simulated tracks, generated particles, and jets clustered from the generated particles. The various objects are matched in order to reconstruct the provenance of the various hits. Samples are produced from the standard CMS format "AODSIM" plus a series of low-level tracker-related collections that allow the extraction of the tracker hits.

Dataset name	Description	Number of events	Number of files
QCD300to600	QCD, flat pT hat spectrum, 300 < pT hat < 600 GeV	1497600	2496
QCD400to600	QCD, flat pT hat spectrum, 400 < pT hat < 600 GeV	1989000	3315
QCD600to3000	QCD, flat pT hat spectrum, 600 < pT hat < 3000 GeV	2974800	4959
tbar	tbar, fully hadronic decays, pT of the top/antitop greater than 400 GeV	2969109	4055

Related datasets

QCD300to600_Run1_8TeV was derived from:
Tracker-hit-enriched 300 to 600 bin of QCD, Pt-15to3000, TuneZ2star, Flat, 8TeV, pythia6

QCD400to600_Run1_8TeV was derived from:
Tracker-hit-enriched 400 to 600 bin of QCD, Pt-15to3000, TuneZ2star, Flat, 8TeV, pythia6

QCD600to3000_Run1_8TeV was derived from:
Tracker-hit-enriched 600 to 3000 bin of QCD, Pt-15to3000, TuneZ2star, Flat, 8TeV, pythia6

tbar_Run1_8TeV was derived from:
Tracker-hit-enriched TJets, HadronicMGDecays, 8TeV, madgraph

Dataset characteristics

9430509 events, 14825 files, 12.8 TB in total.

Dataset semantics

Variable	Type	Description
hit_global_x	std::vector<float>	global x position of the RecHit
hit_global_y	std::vector<float>	global y position of the RecHit
hit_global_z	std::vector<float>	global z position of the RecHit
hit_local_x	std::vector<float>	x pos. of the hit in the local sensor coordinate
hit_local_y	std::vector<float>	y pos. of the hit in the local sensor coordinate
hit_local_x_error	std::vector<float>	x error in the local sensor coordinate
hit_local_y_error	std::vector<float>	y error in the local sensor coordinate
hit_sub_det	std::vector<unsigned int>	subdetector generating the hit [1 PixelBarrel, 2 PixelEndcap, 3 TIB, 4 TID, 5 TOR, 6 TEC]
hit_layer	std::vector<unsigned int>	layer/disk of the subdetector generating the hit
hit_type	std::vector<unsigned int>	Type of strip hit [0 Pixel hit, 1 rphiRecHit, 2 stereoRecHit, 3 rphiRecHitUnmatched, 4 stereoRecHitUnmatched]
hit_simtrack_id	std::vector<int>	ID number of the sim track matched to the hit
hit_simtrack_index	std::vector<unsigned int>	Index of the sim track matched to the hit
hit_simtrack_match	std::vector<bool>	is the hit matched to a sim track?
hit_genparticle_id	std::vector<unsigned int>	Index of the gen particle matched to the hit
hit_pdgid	std::vector<int>	PDG ID of the gen particle matched to the hit
hit_recotrack_id	std::vector<unsigned int>	Index of the reco track matched to the hit
hit_recotrack_match	std::vector<bool>	is the hit matched to a reco track?
hit_genparticle_match	std::vector<bool>	is the hit matched to a gen particle?
hit_genjet_id	std::vector<unsigned int>	Index of the gen jet matched to the hit

Dataset variables coming with detailed semantics description

Interoperable: accompanying examples

opendata CERN Search Help About

Analysis of the di-muon spectrum using data from the CMS detector taken in 2012

Wunsch, Stefan

Cite as: Wunsch, Stefan; (2019). Analysis of the di-muon spectrum using data from the CMS detector taken in 2012. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.AAR1.AN2Q

Selection Analysis Results Data

Description

This analysis takes data from the CMS experiment recorded in 2012 during Run B and C and extracts the di-muon spectrum. The di-muon spectrum is computed from the data by calculating the invariant mass of muon pairs with opposite charge. In the resulting plot, you are able to rediscover particle resonances in a wide energy range from the J/ψ meson at about 548 MeV up to the Z boson at about 91 GeV.

The analysis code opens an interactive plot, which allows to zoom and navigate in the spectrum. Note that the bump at 30 GeV is not a resonance but an effect of the data taking due to the used trigger. The technical description of the dataset can be found in the respective record linked below.

The result of this analysis can be compared with an official result of the CMS collaboration using data taken in 2010, see the plots below:

Dimuon spectrum analysis example

opendata CERN Search Help About

Analysis of Higgs boson decays to two tau leptons using data and simulation of events at the CMS detector from 2012

Wunsch, Stefan

Cite as: Wunsch, Stefan; (2019). Analysis of Higgs boson decays to two tau leptons using data and simulation of events at the CMS detector from 2012. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.GV20.PR5T

Selection Analysis Results Data

Description

This analysis uses data and simulation of events at the CMS experiment from 2012 with the goal to study decays of a Higgs boson into two tau leptons in the final state of a muon lepton and a hadronically decayed tau lepton. The analysis follows loosely the setup of the official CMS analysis published in 2014.

The purpose of the original CMS analysis was to establish the existence of the Higgs boson decaying into two tau leptons. Since performing this analysis properly with full consideration of all systematic uncertainties is an enormously complex task, we reduce this analysis to the qualitative study of the kinematics and properties of such events without a statistical analysis. However, as you can explore in this record, already such a reduced analysis is complex and requires extensive physics knowledge, which makes this a perfect first look into the procedures required to claim the evidence or existence of a new particle.

Two example results produced by this analysis can be seen below. The plots show the data recorded by the detector compared to the estimation of the contributing processes, which are explained in the following. The analysis has implemented the visualization of 34 such observables.

$H \rightarrow \tau\tau$ analysis example

Reusable: how were the data generated?

Simulated dataset BulkGravTohhTohbbbbb_narrow_M-4500_13TeV-madgraph in MINIAODSIM format for 2016 collision data

/BulkGravTohhTohbbbbb_narrow_M-4500_13TeV-madgraph/RunII/Summer16/MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TracheIV_v6-v1/MINIAODSIM, CMS Collaboration

Cite as: CMS Collaboration (2019). Simulated dataset BulkGravTohhTohbbbbb_narrow_M-4500_13TeV-madgraph in MINIAODSIM format for 2016 collision data. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.7N4X.Z7FA

[Dataset](#) [Simulated](#) [Exotica](#) [Gravitons](#) [CMS](#) [13TeV](#) [CERN-LHC](#)

How were these data generated?

These data were generated in several steps (see also [CMS Monte Carlo production overview](#)):

Step LHE

Release: CMSSW_7_1_16

Output dataset: /BulkGravTohhTohbbbbb_narrow_M-4500_13TeV-madgraph/RunII/Winter15wmLHE-MCRUN2_71_V1-v1/LHE

Note: To get the exact generator parameters, please see [Finding the generator parameters](#).

Step SIM

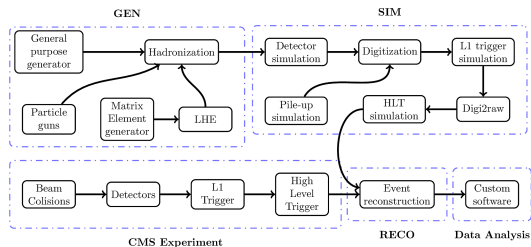
Release: CMSSW_7_1_20

Configuration file for SIM ([link](#))

Output dataset: /BulkGravTohhTohbbbbb_narrow_M-4500_13TeV-madgraph/RunII/Summer15GS-MCRUN2_71_V1-v1/GEN-SIM

Step HLT RECO

Release: CMSSW_8_0_71



- ▶ full capture of data generation steps
- ▶ full capture of compute environments
- ▶ full capture of configuration files
- ▶ full capture of production scripts

Data records come with full provenance information

Reusable: can we reprocess published data samples?

SingleElectron primary dataset sample in RAW format from RunA of 2011 (from /SingleElectron/Run2011A-v1/RAW)

/SingleElectron/Run2011A-v1/RAW, CMS collaboration

Cite as: CMS collaboration (2019). SingleElectron primary dataset sample in RAW format from RunA of 2011 (from /SingleElectron/Run2011A-v1/RAW). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.6O84.WLN8

Dataset Collision CMS 7TeV CERN-LHC

Description

A sample from SingleElectron primary dataset in RAW format from RunA of 2011. Run range [161224,163286].

This dataset contains selected runs from 2011 RunA. The list of validated lumi sections, which must be applied to all analyses on events reconstructed from these data, can be found in

[CMS list of validated runs Cert_160404-180252_7TeV_ReRecoNov08_Collisions11_JSON.txt](#)

Dataset characteristics

2064298 events. **116** files. **424.3 GB** in total.

How can you use these data?

These data are in RAW format and not directly usable in analysis. The reconstructed data reprocessed from these RAW data are included in the data of [this record](#). The reconstruction step can be repeated with the configuration file below and the resulting AOD has been confirmed to be identical with the original one with comparison code available in [Validation code to plot basic physics objects from AOD](#)

RAW

SingleElectron primary dataset in AOD format from RunA of 2011 (/SingleElectron/Run2011A-12Oct2013-v1/AOD)

/SingleElectron/Run2011A-12Oct2013-v1/AOD, CMS collaboration

Cite as: CMS collaboration (2016). SingleElectron primary dataset in AOD format from RunA of 2011 (/SingleElectron/Run2011A-12Oct2013-v1/AOD). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.P87Z.TXTV

Dataset Collision CMS 7TeV CERN-LHC

Description

SingleElectron primary dataset in AOD format from RunA of 2011. Run period from run number 160404 to 173692.

This dataset contains all runs from 2011 RunA. The list of validated runs, which must be applied to all analyses, can be found in

[CMS list of validated runs Cert_160404-180252_7TeV_ReRecoNov08_Collisions11_JSON.txt](#)

Dataset characteristics

41709195 events. **1542** files. **5.8 TB** in total.

How were these data selected?

Events stored in this primary dataset were selected because of the presence of at least one high-energy [electron](#) in the [event](#).

Data taking / HLT

The collision data were assigned to different RAW datasets using the following [HLT configuration](#).

Data processing / RECO

This primary AOD dataset was processed from the RAW dataset by the following step:

Step: RECO

Release: CMSSW_5_3_12_patch1

Global tag: FT_R_53_LVS:All

[Configuration file for RECO step reco_2011A_SingleElectron](#)

AOD

Reprocessing workflow: computational recipe

3. Workflow

The workflow can be logically divided into several parts:

0. Upload all files.

Some files cannot be generated at run time and need to be uploaded.

```
inputs:
files:
- src/PhysicsObjectsHistos.cc
- BuildFile.xml
- demoanalyzer_cfg.py
```

1. Fix the CMS SW environment variables manually.

First, we have to set up the environment variables accordingly for the [CMS SW](#). Although this is done in the docker image, reana overrides them and they need to be reset. This is done by invoking the [cms entrypoint.sh](#) script commands.

See also [this issue](#).

```
$ source /opt/cms/cmsset_default.sh
$ scramv1 project CMSSW CMSSW_5_3_32
$ cd CMSSW_5_3_32/src
$ eval `scramv1 runtime -sh`
```

2. Create the specific CMS path.

CMS specific data analysis framework requires two directory levels. See also [this issue](#).

```
$ mkdir Reconstruction && cd Reconstruction
$ mkdir Validation && cd Validation
```

3. Create the reconstruction file.

See also [this repo](#).

```
$ cmsDriver.py reco -s RAM2DIGI,L1Reco,RECO,USER:EventFilter/HcalRawToDigi/hcallaserbhbheffilter2012_cf
```

4. Adjust the reconstruction file to the specific data file.

Although generated using parameters, the reconstruction file still requires changes.

```
$ sed -i 's/from Configuration.AiCa.GlobalTag import GlobalTag/process.GlobalTag.connect = cms.string("
$ sed -i 's/# Other statements/from Configuration.AiCa.GlobalTag import GlobalTag/g' reco_cmsdriver.py
$ sed -i 's/process.GlobalTag = GlobalTag(process.GlobalTag, 'FT_53_LV5_ANI::All', '')/process.GlobalTag
```

5. Link the CVMFS files.

The `-l` commands are explicitly needed to make sure that the `cms-opendata-conddb.cern.ch` directory has actually expanded in the image, according to [this guide](#). See also [this issue](#).

```
$ ln -sf /cvmfs/cms-opendata-conddb.cern.ch/FT_53_LV5_ANI_RUNA_FT_53_LV5_ANI
$ ln -sf /cvmfs/cms-opendata-conddb.cern.ch/FT_53_LV5_ANI_RUNA.db_FT_53_LV5_ANI_RUNA.db
$ ls -l
$ ls -l /cvmfs/
```

6. Run the reconstruction.

At this point all environment variables and files should be proper.

```
$ cmsRun reco_cmsdriver.py
```

7. Adjust project structure for validation

Copy the required files for the next steps.

```
$ mkdir src
$ scp ../../../../src/PhysicsObjectsHistos.cc ./src
$ scp ../../../../BuildFile.xml .
$ scp ../../../../demoanalyzer_cfg.py .
```

8. Run CMS scram command to fix libraries.

Most importantly, the `BuildFile.xml` has to be inside the directory where the `scram` command is executed.

```
$ scram b
```

9. Run the validation file.

See also [this repo](#)

```
$ cmsRun demoanalyzer_cfg.py
```

Workflow steps to run CMS reconstruction in CMSSW environment

Reprocessing CMS datasets on REANA

dataset=Jet
year=2011A

1 input
parameters

```
$ cms-reco --create-workflow  
Created 'cms-reco-SingleElectron-2011' directory.  
$ cd cms-reco-SingleElectron-2011  
$ reana-client run
```

2 workflow
factory



```
reana:
  name: cms-reco-SingleElectron-2011
  namespace: cms-reco-SingleElectron-2011
  workflow:
    name: cms-reco-SingleElectron-2011
    namespace: cms-reco-SingleElectron-2011
    version: 1.0.0
    description: cms-reco-SingleElectron-2011
    inputs:
      dataset: Jet
      year: 2011A
    outputs:
      dataset: Jet
      year: 2011A
    steps:
      - name: cms-reco-SingleElectron-2011
        namespace: cms-reco-SingleElectron-2011
        version: 1.0.0
        description: cms-reco-SingleElectron-2011
        inputs:
          dataset: Jet
          year: 2011A
        outputs:
          dataset: Jet
          year: 2011A
        command: cms-reco --create-workflow --dataset Jet --year 2011A --reana-client run
```

3

reana.yaml

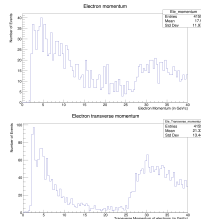


reana

4 run by REANA platform



5 serving open data files



6 output
histograms

Parametrised workflow runnable on REANA reproducible analysis platform

Latest technology news

The screenshot shows the homepage of the opendata CERN website. At the top left, the logo reads "opendata CERN". At the top right, there are links for "Help" and "About". Below the header, a yellow banner contains an "Important notice" stating that the current site is a development server and should be used for testing only, with a link to the production site "opendata.cern.ch".

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, search examples are provided: "collision datasets, keywords:education, emesa:ZNY".

There are two columns of links:

- Explore**
 - [datasets](#)
 - [software](#)
 - [environments](#)
 - [documentation](#)
- Focus on**
 - [ATLAS](#)
 - [ALICE](#)
 - [CMS](#)
 - [LHCb](#)
 - [OPERA](#)
 - [Data Science](#)

At the bottom center, there is a "Get started" button with a dropdown arrow.

- ▶ **March 2021:** Major upgrade of the underlying digital repository software framework (Invenio 3.4, Elasticsearch 7, Angular → React, Bootstrap → SemanticUI, ...)

Future plans

- ▶ **December 2020:** A common statement on the open data policy by CERN management and ATLAS, ALICE, CMS, LHCb and TOTEM experiments.

<https://opendata.cern.ch/docs/cern-open-data-policy-for-lhc-experiments>

- ▶ Prepare for forthcoming increase in open data publishing.
- ▶ Introduce flexible hot/cold disk/tape storage solution. Part of dataset files on disk, part on tapes.
- ▶ Simplify ingestion and exposure of experiment datasets (Rucio, Dirac).
- ▶ Further automatise provenance testing and usage examples.

CERN announces new open data policy in support of open science

A new open data policy for scientific experiments at the Large Hadron Collider (LHC) will make scientific research more reproducible, accessible, and collaborative

11 DECEMBER, 2020



Data storage solutions at the CERN data centre (Image: CERN)

Geneva, 11 December 2020. The four main LHC collaborations (ALICE, ATLAS, CMS and LHCb) have unanimously endorsed a new open data policy for scientific experiments at the Large Hadron Collider (LHC), which was presented to the CERN Council today. The policy commits to publicly releasing so-called level 3 scientific data, the type required to make scientific studies, collected by the LHC experiments. Data will start to be released approximately five years after collection, and the aim is for the full dataset to be publicly available by the close of the experiment concerned. The policy addresses the growing movement of open science, which aims to make scientific research more reproducible, accessible, and collaborative.

The level 3 data released can contribute to scientific research in particle physics, as well as research in the field of scientific computing, for example to improve reconstruction or analysis methods based on machine learning techniques, an approach that requires rich data sets for training and validation.

An ecosystem of sister data repositories

HEPData widens publication-level data scope

HEPData Search HEPData

Reconstruction and identification of boosted di- τ systems in a search for Higgs boson pairs using 13 TeV proton-proton collision data in ATLAS

The ATLAS collaboration

Abstract (info abstract)

Table 1

Stage of selection	Pre-selection	Di- τ selection	Large- J_T jet selection	Signal region
1000	0.24551	0.051792	0.012219	0.0020
1190	0.03949	0.078988	0.032029	0.0089
1280	0.40346	0.162247	0.055908	0.0183
1480	0.49898	0.136511	0.089954	0.0355
2000	0.059782	0.139633	0.11138	0.0350

Visualize

Sum errors Log Scale (X)

ATLAS collaboration <https://www.hepdata.net/record/ins1809175>

HEPData provides interactive interface to explore and download publication-level data behind plots and tables

HEPData Search HEPData

Search for bottom-squark pair production with the ATLAS detector in final states containing Higgs bosons, b-jets and missing transverse momentum

The ATLAS collaboration

Abstract

Table 1

cmenergies	observables	phrases	reactions
13000	ATLAS	Production	PP -> BOTTOM BOTTOM

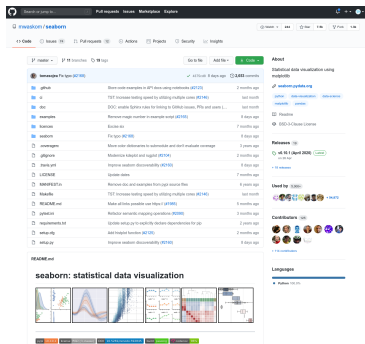
Visualize

Sum errors Log Scale (X)

ATLAS collaboration <https://www.hepdata.net/record/ins1748602>

... now starting to handle more data types: likelihoods!

Zenodo preserves research software

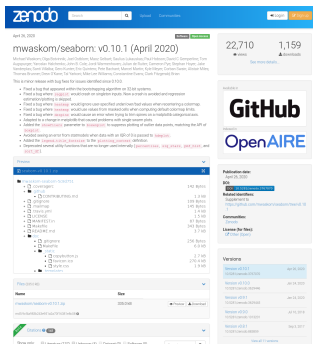


Latest release

v0.10.1

dd40fd6

DOI 10.5281/zenodo.3767070



<https://guides.github.com/activities/citable-code>

GitHub ↔ Zenodo bridge to automatically preserve releases

Zenodo hosts Machine Learning community data

Sample with tracker hit information for tracking algorithm ML studies TTbar_13TeV_PU50_PixelSeeds

Di Florio, Adriano; Pantaleo, Felice; Pierini, Maurizio; (2019). Sample with tracker hit information for tracking algorithm ML studies TTbar_13TeV_PU50_PixelSeeds. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.NLUN.TQHD

Dataset **Dataset** **Dataset** **CMS** **CERN-CD**

Parent Dataset: TTbar13TeV_PU50_PixelSeeds generated by LHC Phase2 studies

Description

This dataset consists of a collection of pixel doublet seeds, i.e. the hit pairs that could belong to the same particle flying through the CMS Silicon Pixel Detector. These can be used in ML studies of particle tracking algorithms. Particle tracking is the process of clustering the recorded hits into groups of points arranged along an helix.

One of the first steps of the track finding workflow is the creation of track seeds, i.e. compatible pairs of hits from different detector layers, that are subsequently fed to higher level pattern recognition steps. The compatibility between two hits is evaluated only on the basis of geometrical considerations, such as cuts in ϕ and r , and these doublets define the building blocks for further tracks. However the set of compatible hit pairs is highly affected by combinatorial background resulting in the next steps of the tracking algorithm to process a significant fraction of fake doublets. For each event an $O(10^6)$ doublets are produced while only an $O(10^3)$ are genuine resulting in a fake ratio of $O(10^3)$.

Each doublet in this dataset is characterised by a set of features, such as its coordinates and the charge released in their Pixel Detector, and the pixel cluster shape, projected on 2D histogram. A possible way of reducing the huge combinatorial background is using Machine Learning and Deep Learning techniques to check the compatibility between two hits. Indeed the task of fake rejection can be seen as a typical classification problem for which networks and MVA methods have been widely proven to provide reliable results. The dataset provided is intended to be used to explore this techniques.

Di Florio et al <http://opendata.cern.ch/record/12320>

Machine-learning related datasets on the CERN Open Data portal

CMS 2011A Simulation | Pythia 6 QCD 300-470 | pT > 375 GeV | MOD HDF5 Format

71 views 765 downloads

OpenAIRE

Publication date: August 6, 2019

DOI: 10.52000/ZENODO.3341498

Keywords: CMS, Pythia 6, QCD, Simulation, Particle Physics, High Energy Physics

Related identifiers: arXiv:1907.07381v1, 10.52000/ZENODO.3341498

License for this file: CC BY-NC-ND 4.0 International

There is an associated dataset of data generated by the CMS detector available on Zenodo:

- CMS 2011A_sim_v1 - 375 GeV

Name	Size	Download
ZENODO_3341498_1.compressed.tif	102.4 MiB	Download
ml-hits-tracker-hit-information-for-tracking-algorithm-ml-studies-ttbar-13tev-50-pu-pixel-seeds	102.4 MiB	Download
ZENODO_3341498_2.compressed.tif	102.4 MiB	Download
ml-hits-tracker-hit-information-for-tracking-algorithm-ml-studies-ttbar-13tev-50-pu-pixel-seeds	102.4 MiB	Download
ZENODO_3341498_3.compressed.tif	102.4 MiB	Download
ml-hits-tracker-hit-information-for-tracking-algorithm-ml-studies-ttbar-13tev-50-pu-pixel-seeds	102.4 MiB	Download
ZENODO_3341498_4.compressed.tif	102.4 MiB	Download
ml-hits-tracker-hit-information-for-tracking-algorithm-ml-studies-ttbar-13tev-50-pu-pixel-seeds	102.4 MiB	Download

File history

Version 1.000 created on 2019-08-06

File of related records: ml-hits-tracker-hit-information-for-tracking-algorithm-ml-studies-ttbar-13tev-50-pu-pixel-seeds

Share: Copy as: arXiv, PDF, Markdown, Dublin Metadata, BibTeX, Previews & Thumbnails, DOI, CMS 2011A Simulation | Pythia 6 QCD 300-470 | pT > 375 GeV | MOD HDF5 Format | Zenodo: 48, Datacite: 48, Zenodo: 10, MyData on Zenodo: 3341498

Komiske et al <https://zenodo.org/record/3341498>

Machine-learning related datasets on Zenodo

Preserving “restricted” analysis knowledge



Home What is CAP? Get Started Integrations Documentation Log In

CERN Analysis Preservation

capture, preserve and reuse physics analyses



Capture



Collect and preserve elements
needed to understand and
rerun your analyses

Collaborate



Share your analysis and
components with other users,
your collaboration or group

Reuse



Run containerized workflows
and easily reuse analysis
components

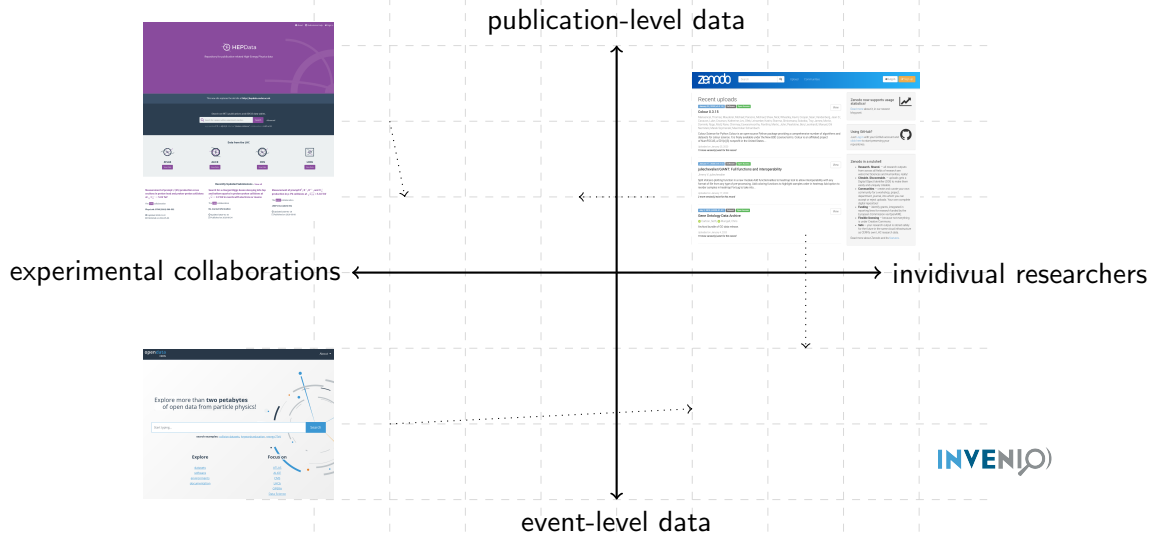
The screenshot shows the CERN Analysis Preservation web interface. On the left, there are filter sections for TYPE, COLLISION SYSTEM, CADI STATUS, and ACCELERATOR PARAMETERS. The main area displays a list of results with titles like "Light by Light scattering with PbPb 5.02 TeV data" and "Base-Electron correlations in various collision systems and energies". Below the results, a terminal window shows the following commands and output:

```
$ pip install csp-client
$ export CAP_SERVER_URL=https://analysispreservation.cern.ch/
$ export CAP_ACCESS_TOKEN=<your generated access token from server>
$ csp-client files upload <file path> --pid/-p <existing pid>
$ csp-client files upload file.json -p #9b593c498874ec8bcafc8934c439a7
File uploaded successfully.
```

<https://analysispreservation.cern.ch>

CERN Analysis Preservation framework for collaboration-restricted data.
Following the same FAIR principles (FAIR \neq open!)

A family of digital repositories in movement



INVENIO