

Resources & Tools to go further with Open Data

JULIE HOGAN

8.01.24

Analysis in any format: AOD (Run 1)

AOD data provides the most “low-level” information, while still containing user-level physics objects like jets, electrons, muons, photons. <https://opendata.cern.ch/docs/cms-getting-started-aod>

Three workshops feature Run 1 AOD analysis content:

- 2020: <https://cms-opendata-workshop.github.io/2020-09-30-cms-open-data-workshop-for-theorists/>
- 2021: <https://cms-opendata-workshop.github.io/2021-07-19-cms-open-data-workshop/>
- 2022: <https://cms-opendata-workshop.github.io/2022-08-01-cms-open-data-workshop/>

The Open Data Team has created a suite of “EDAnalyzers” that can process AOD into a ROOT tree format similar to NanoAOD

- 2011 & 2012 branches: <https://opendata.web.cern.ch/record/12501>
- Some common ID, iso, corrections applied in the code. Analysis examples in the workshops!

Another “Run 1 NanoAOD” producer available: <https://opendata.web.cern.ch/record/12505>

- Coupled with example analyses that show two different analysis techniques: <https://opendata.web.cern.ch/record/12506>

Analysis in any format: MiniAOD (Run 2)

MiniAOD data provides a comprehensive set of user-level physics objects, with important lower-level items like tracks & candidates. <https://opendata.cern.ch/docs/cms-getting-started-miniaod>

Two workshops feature Run 2 MiniAOD analysis content:

- 2022: <https://cms-opendata-workshop.github.io/2022-08-01-cms-open-data-workshop/>
- 2023: <https://cms-opendata-workshop.github.io/2023-07-11-cms-open-data-workshop/>

The Open Data Team has created a suite of “EDAnalyzers” that can process AOD into a ROOT tree format similar to NanoAOD

- 2015 branch: <https://opendata.web.cern.ch/record/12502>
- Some common ID, iso, corrections applied in the code
- Adds jet substructure information for “FatJets”
- Analysis example in the 2022 and 2023 workshops

Analysis in any format: NanoAOD (Run 2)

NanoAOD data provides commonly used high-level physics object information suitable for many analyses. <https://opendata.cern.ch/docs/cms-getting-started-nanoaod>

This workshop is the first to feature NanoAOD!

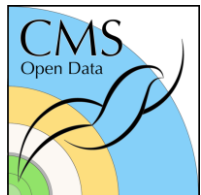
- 2024: <https://cms-opendata-workshop.github.io/2024-07-29-CERN/>

No “processing” framework is needed, we can jump straight to analysis examples :D

- See the analysis example in the workshop – python tools (uproot, awkward)
- One tool developed within CMS can be helpful for pre-filtering NanoAOD for validated runs & other cuts: <https://opendata.web.cern.ch/record/12507>
- Run 1 derived NanoAOD examples (loops, RDataFrame): <https://opendata.web.cern.ch/record/12506>
- Analysis Grand Challenge (COFFEA): <https://agc.readthedocs.io/en/latest/index.html>

Example producer tool for PFNano is a template for how to add more things to NanoAOD:

- <https://opendata.cern.ch/record/12504>



CMS Open Data Guide

The Open Data Portal hosts the data and many “technical” help pages

- Managed by CERN IT

The CMS Open Data Guide will be our home for analysis information

- Managed by us!

The goal is for the Guide to contain necessary common analysis information for AOD, MiniAOD, NanoAOD format

CMS Open Data Guide



Warning

This guide is under construction

Welcome to the guide for CMS open data. This guide is brought to you by the CMS open data group, on a best-effort basis. All software and instructions are provided "as is", without warranty of any kind. This is ongoing work and we appreciate your feedback and/or your help building this guide.

Goal to get the Guide up-to-date for 2016 this year!!

CMS Open Data Guide

Home

CMS Open Data



Computing Tools



CMSSW



Analysis



Data and Simulation



Selection



Luminosity

Background Modelling

Systematics



Statistical Interpretation

Physics Object info

CMS Open Data Guide

Triggers

Physics Objects ▼

Physics Objects

Common Tools

Muons

Electrons

Photons

Jets

Missing ET

Taus

Guide has pages for each of the common physics objects with access example similar to the “Physics Objects” pre-learning. **NanoAOD will be added!**

Accessing Jets in CMS Software

Run 1 Data Run 2 Data

Two examples of EDAnalyzers accessing jet information are available in the Physics Object Extractor Tool (POET):

- [JetAnalyzer](#) accessing jets from the `PFJetCollection`
- [PatJetAnalyzer](#) accessing jets from `std::vector<pat::Jet>` collection using "Physics Analysis Toolkit" (PAT) format in which jets are easier to work with.

The following header files needed for accessing jet information are included:

```
//classes to extract PFJet information
#include "DataFormats/JetReco/interface/PFJet.h"
#include "DataFormats/JetReco/interface/PFJetCollection.h"
#include "DataFormats/BTauReco/interface/JetTag.h"
#include "CondFormats/JetMETObjects/interface/JetCorrectionUncertainty.h"
```

Object corrections

Biggest missing element of the Guide is documenting object corrections

- Jet corrections are documented for Run 1
- Need to find and add lepton correction info for Run 1
- CorrectionLib examples for Run 2, like our Uncertainties lesson
- 2015 corrections are being ported into the common JSON format this summer by a student

Object Uncertainty

Warning

This page is under construction

Implementing JEC in CMS Software

JEC From Text Files

There are several methods available for applying jet energy corrections to reconstructed jets. We have demonstrated a method to read in the corrections from text files and extract the corrections manually for each jet. In order to produce these text files, we have to run [jec_cfg.py](#).

```
isData = False
#if len(sys.argv) > 1: isData = bool(eval(sys.argv[1]))
#print 'Writing JEC text files. isData = ',isData
```



Analysis info

Much of the workshop content will be ported to the Guide for non-physics-object analysis info

Background modeling and Statistical Intepretation pages are already in place!

We **do not** plan to include detailed tutorials or analysis examples in specific frameworks! But may build a page of references to examples in ROOT, uproot/awkward, RDataFrame, Coffea...

Analysis ▼

Data and Simulation >

Selection >

Luminosity >

Background Modelling >

Systematics >

Statistical Interpretation >

Background Modelling Techniques

 Warning

This page is under construction

Table of contents

Simulation

Tight / loose or "Matrix" methods

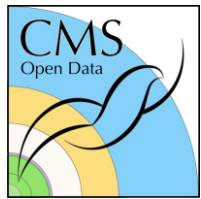
Transfer factors

Alpha-ratio method

ABCD method

Sideband fits

Accurate modeling of SM background processes is essential to most searches and measurements in



Open Data Forum

Get help from us and help others on the Open Data Forum!

opendata

categories ▾ Latest Unread (7) Top Categories

☰ Topic

🚩 Welcome to the CERN Open Data forum!
■ News
Welcome to the CERN Open Data discussion forum! The CERN Open Data Forum manages several petabytes of open data from particle physics. The data is released by the LHC collaborations in periodic batches after a certain em... read more

Library conflict in Simplified Run 2 Analysis lessons from 202...
■ Software tools

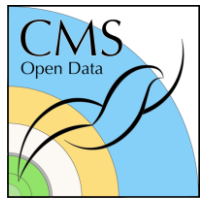
Large data set analysis using only Python 1
■ Containers

Alternatives to Jupyter Notebooks in Python Docker (Windows) 1
■ Containers

opendata
CERN

Help ▾ About ▾

Explore more than **five petabytes**
of open data from particle physics!



Join the effort!

CMS people – we need you!

- EPR available for DPOA tasks, such as contributing to these workshops, the Guide, upcoming releases...
- Reminder for publication policy: <https://cms-docdb.cern.ch/cgi-bin/DocDB/ShowDocument?docid=14372>
 - General guideline is that PubComm needs to be alerted, discuss with them an PC the contents / goal of the publication!
 - Many examples of limited author CMS contributions to papers!

Non-CMS people – we need you!

- We need to build out a suite of NanoAOD example analyses
- Exercise these techniques in your work and provide feedback