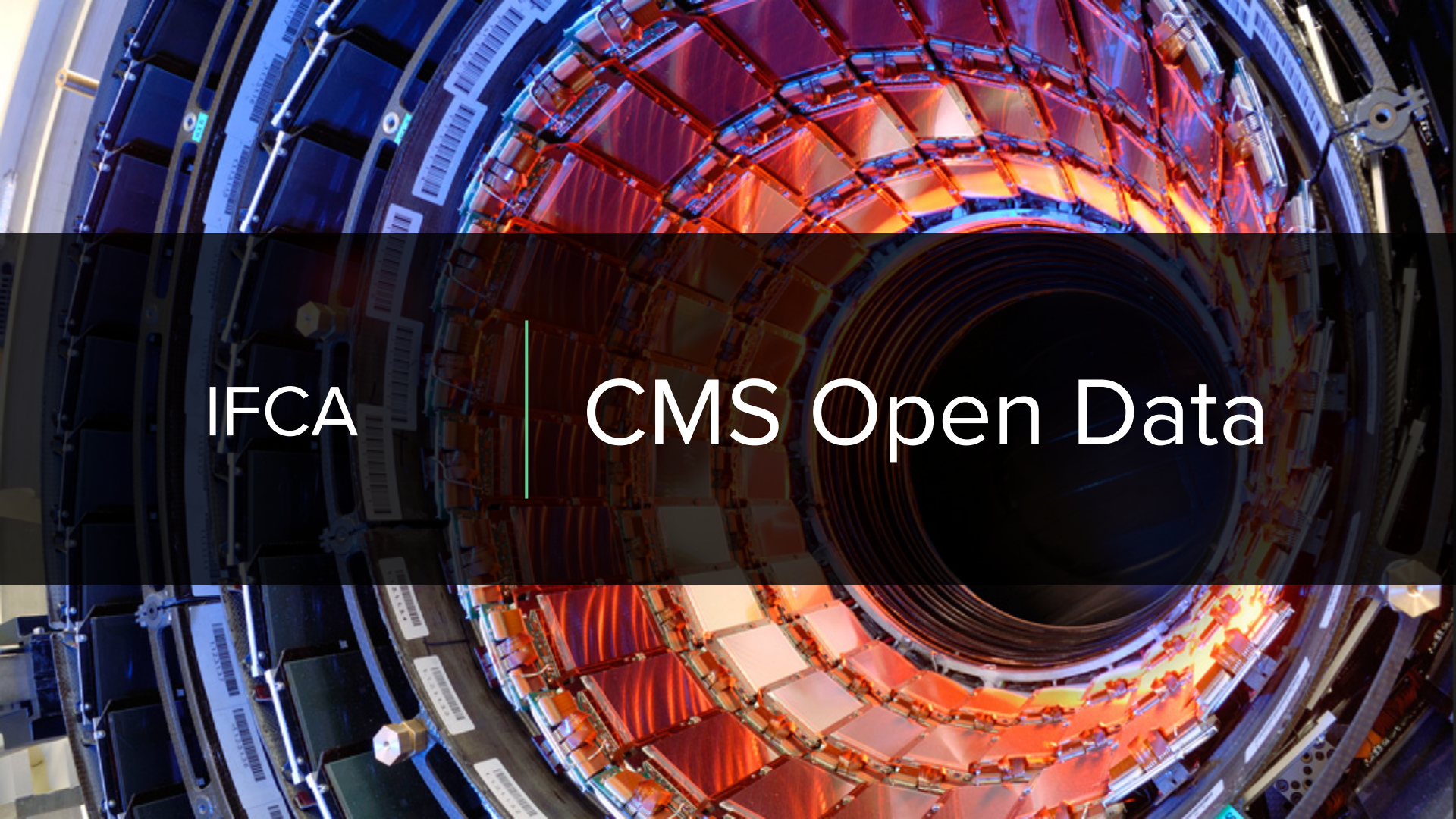

IFCA
CMS Open Data

Table of Contents

- Description of the project
- CMS Open Data Analysis Software
- HEP Ontology

Project's Structure

Generate

Simplified code to analyse CMS events and search for the Z boson.

Conceive

Creation of an Ontology based on CMS Experiment, focused on Z boson.

Update

Improved user interface of the CMS Open Data portal @IFCA

- Add ipython notebook
- Add interface for SPARQL

User's Guide

ANALYSIS SOFTWARE

Destinated to

- High-school students
- Bachelor students

Improvements

- No programming skills needed
- Interactive interface (no need to use the terminal)

Software's structure

3 Classes:

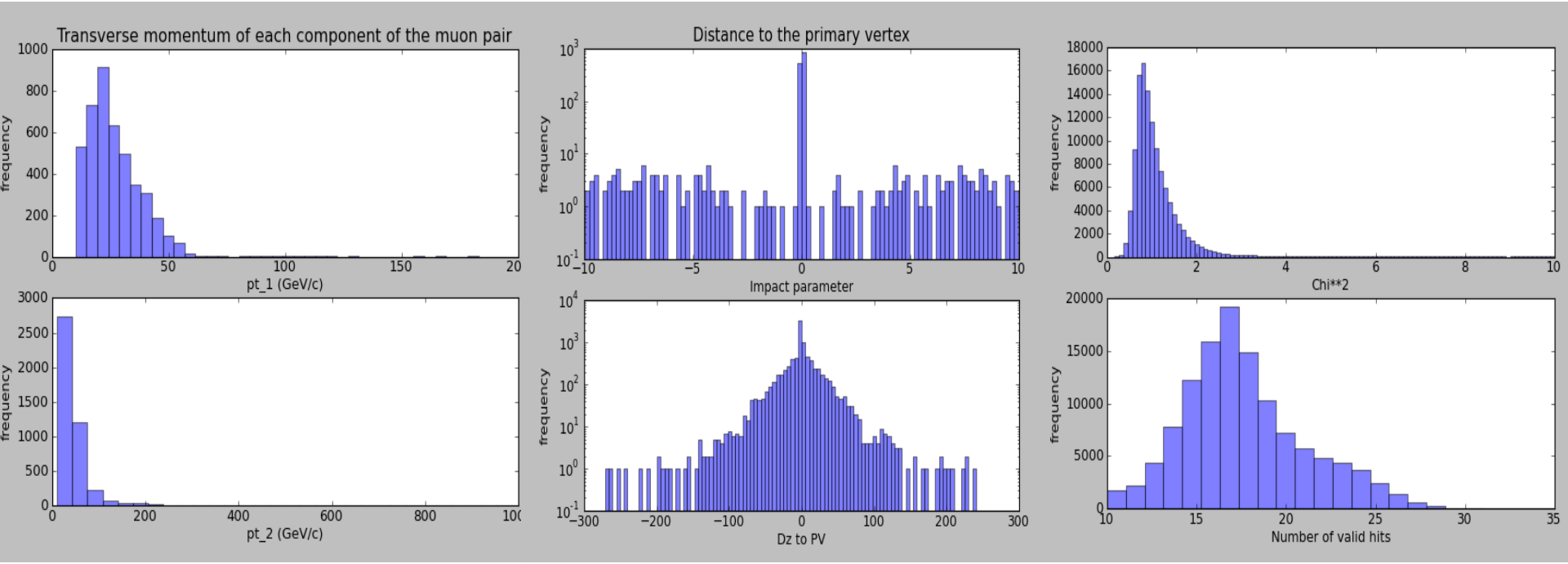
- TwoMuonAnalyzer: analyzes and plots the data
- LeptonPair: paires up the muons and gets their mass and transverse momentum
- CutsConfig: cuts configuration applied to select the good muons

ANALYSIS SOFTWARE: What they will use

Contains:

- The data (ROOT files)
- CutsConfig object
- TwoMuonAnalyzer object: 3 exercises (TwoMuonAnalyzer's functions)

Exercise 1

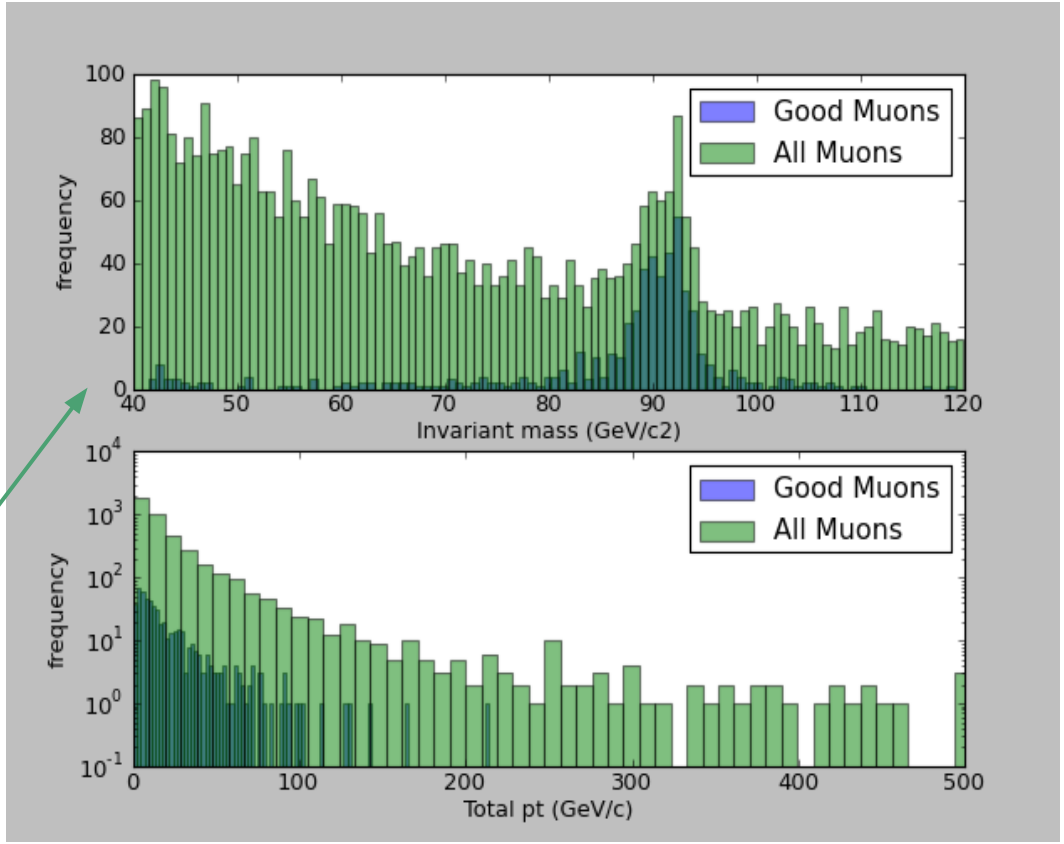

Transverse momentum of each component of the muon pair
frequency
pt_1 (GeV/c)
Distance to the primary vertex
frequency
Impact parameter
frequency
Chi**2
frequency
pt_2 (GeV/c)
frequency
Dz to PV
frequency
Number of valid hits

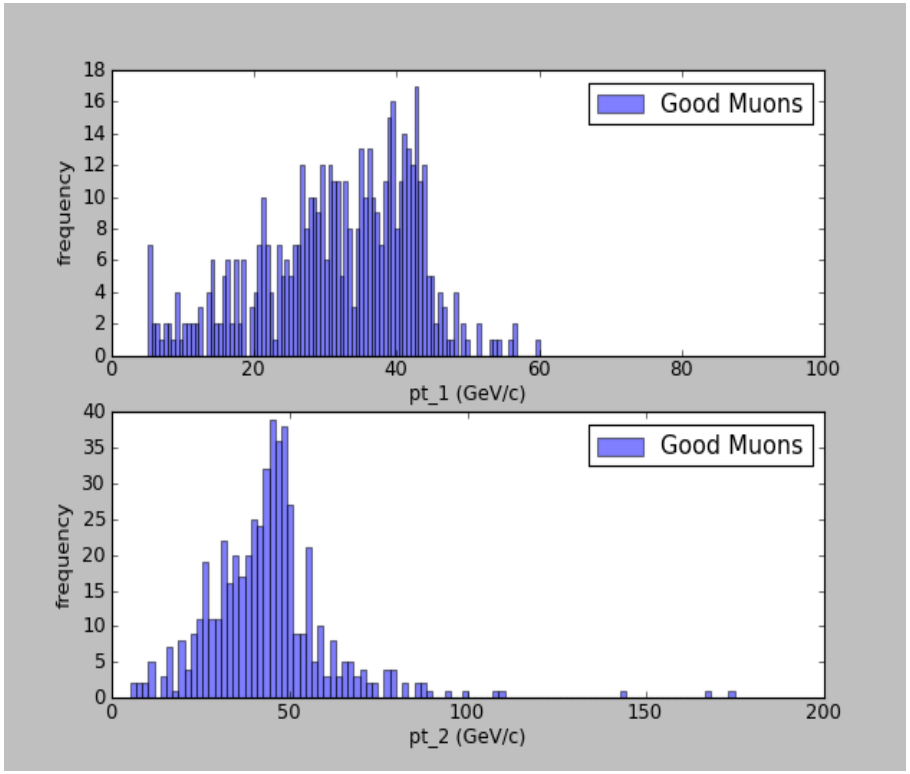
Exercise 2

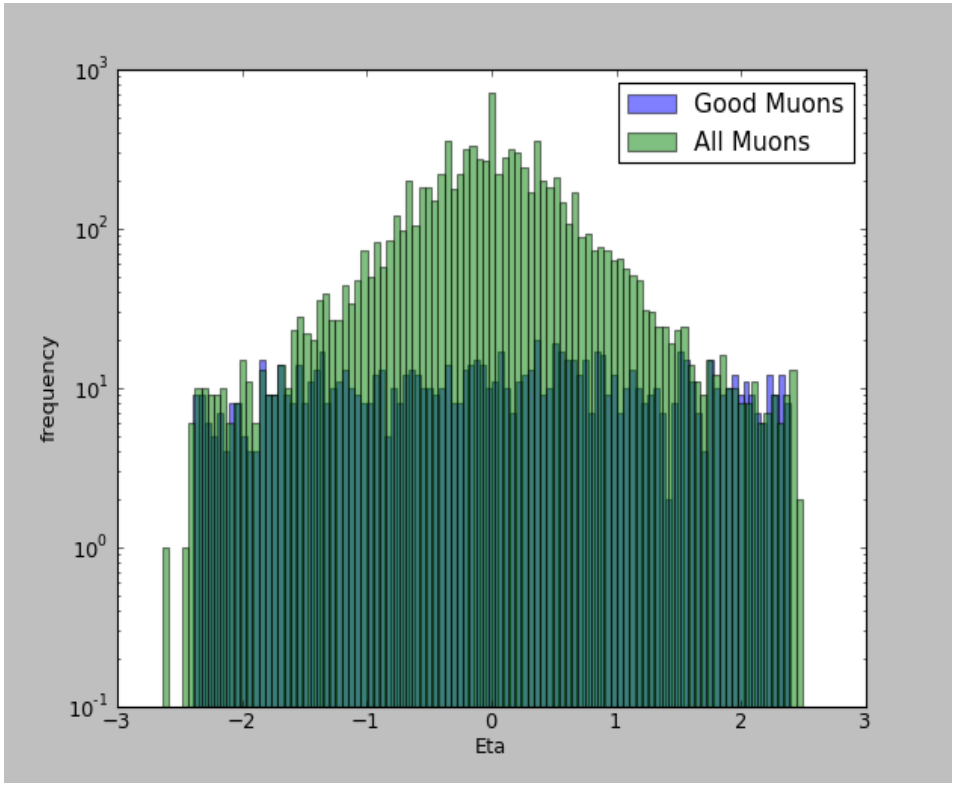
They change this

```
# cutsConfig parameters:

# These are the cuts applied to the muons in order
#to select the good ones (see TwoMuonAnalyzer.py)

pt_min = 5  # Minimum transverse momentum
eta_max = 2.4  # Maximum eta angle
distance = 0.2 # Maximum dz to PV
dB_max = 0.02  # Maximum impact parameter
chi2 = 10  # Maximum chi**2
numValidHits = 10 # Minimum number of valid hits
```

To get the Z boson peak



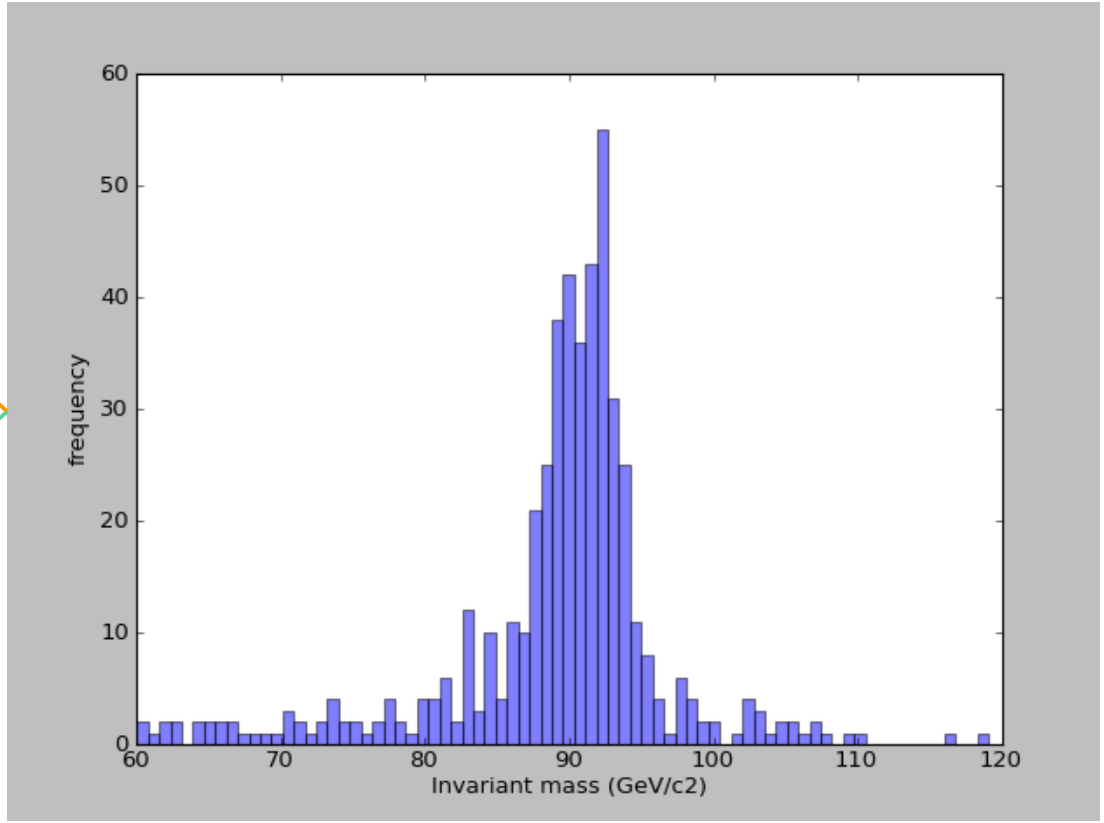

Good Muons
frequency
pt_1 (GeV/c)
Good Muons
frequency
pt_2 (GeV/c)


Good Muons
All Muons
frequency
Eta

Exercise 3: Fit the peak

Gaussian

Breit-Wigner



The problem

Incompatibility of versions

Conflict between VM's python version (2.6) and ipython notebook.

Update of python not possible due to CMS FW Lite version in 2010

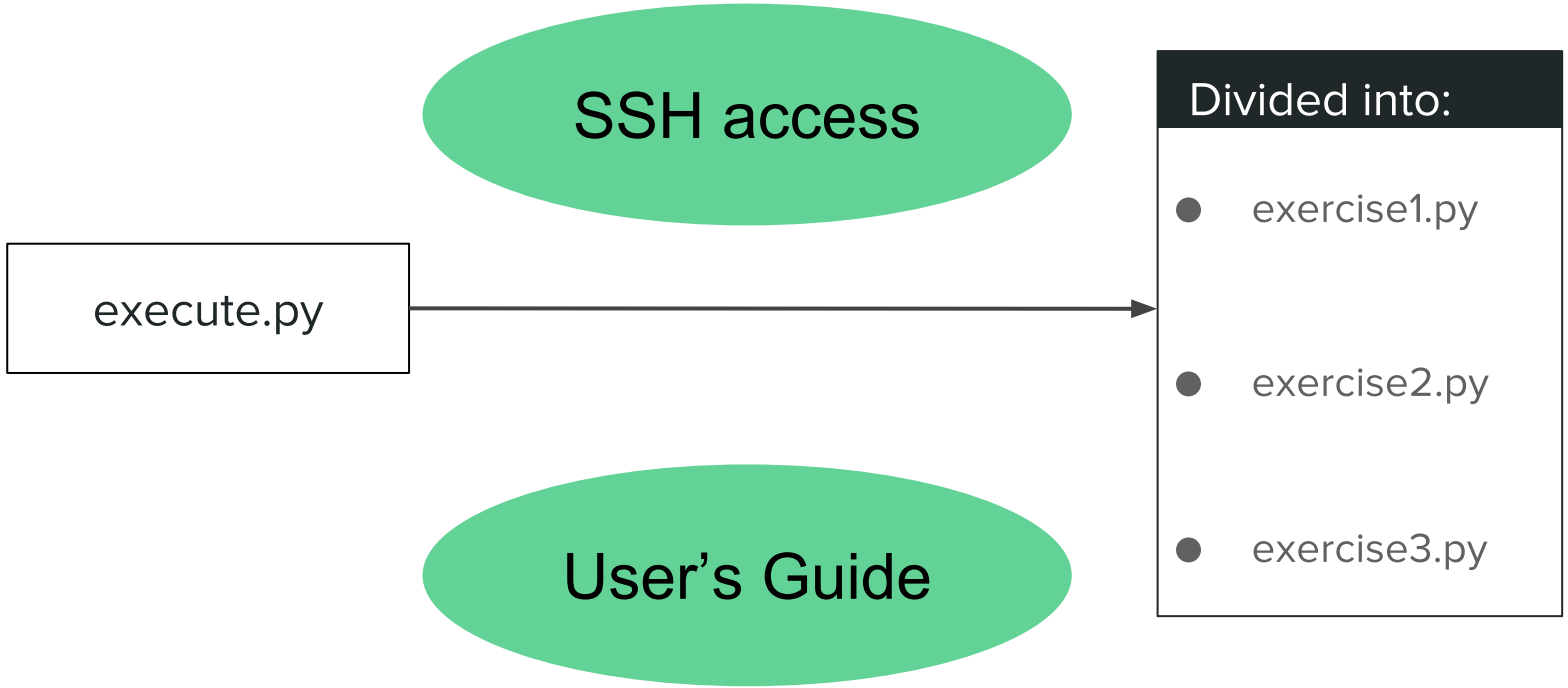
The solution

Make a new format

A new Level 2 data format completely apart from CMS FW Lite is required in public science.

e.g. CSV, JSON, HDF

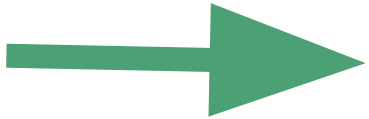
MEANWHILE...


SSH access
execute.py
User's Guide
Divided into:
exercise1.py
exercise2.py
exercise3.py

Need for a Semantic Preservation

HEP Challenges

- ➔ Reproducibility of Scientific Work
- ➔ Common workframe containing general HEP vocabulary
- ➔ Organization of natural and accepted analysis procedures in HEP
- ➔ Everyday semantic map for new students in HEP starting to analyze collisions



Semantic Framework for data analyses containing steps and scientist taking part in the analysis able to be documented, annotated and shared among the scientific HEP comunity

HEP ONTOLOGY

HEP Ontology Definition and Design

➔ Collections of **concepts** and their **relations** based on linked data used in a domain of knowledge with a conceptual vocabulary

➔ Advantages:
- Promote **knowledge sharing**
- **Machine readable**
- Ease data **reuse** and **annotation**
- **scientific workflows**

CMS Open Data Ontology Structure: Main Classes

Standard Model

Includes basic semantic ideas and vocabulary required to explain conceptually the standard model.

- Fundamental Forces
- Lagrangian
- Particles
- Properties

Events

Includes components of Events together with main typical vocabulary

- DataSet
- Physics Objects
- Magnitudes
- Vertex

Analysis

Includes all required parts for analysis and detection of a particle

- CMS Detectors
- Goal Particles
- Candidates Particles
- Restriction and measurements
- Tracks Reconstruction

Software

Collects information and metada corresponding to the software developed for analyzing

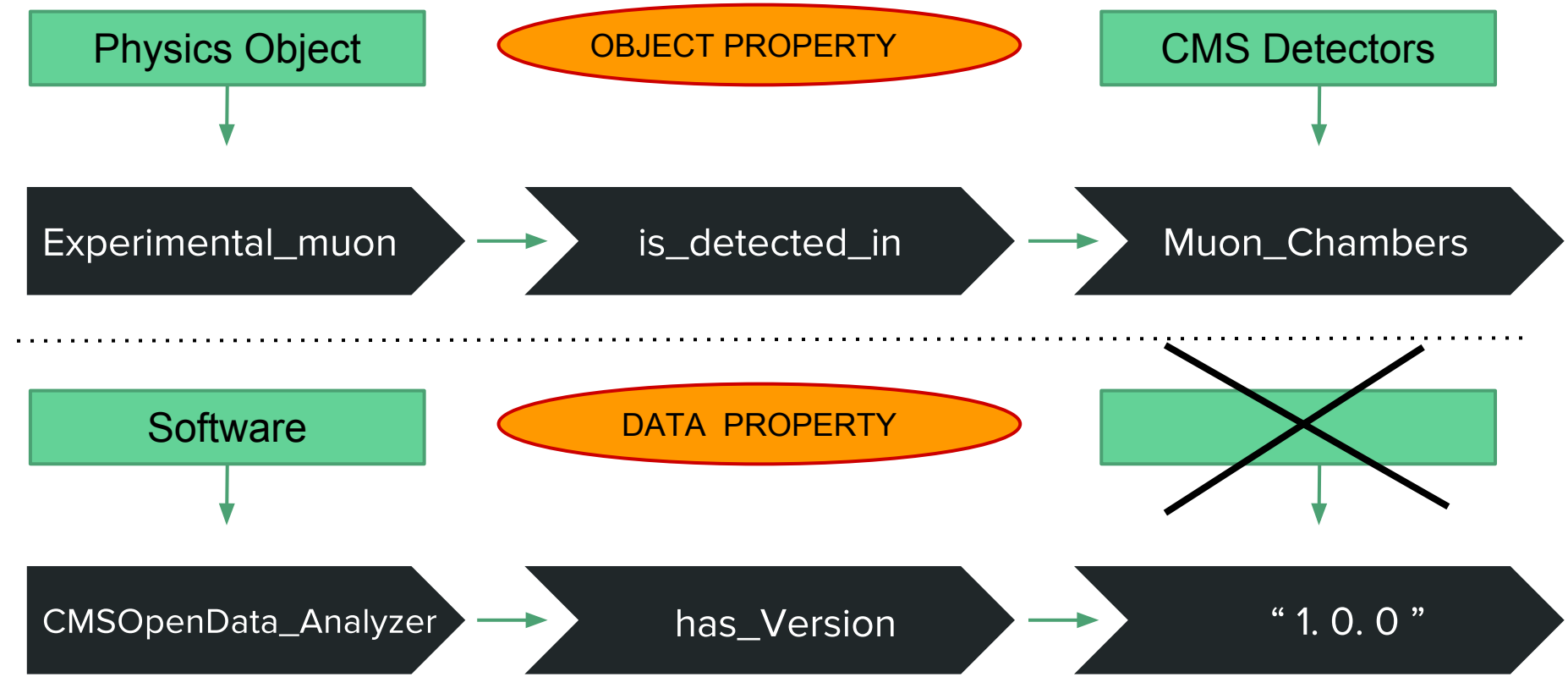
- Execute.py
- Package

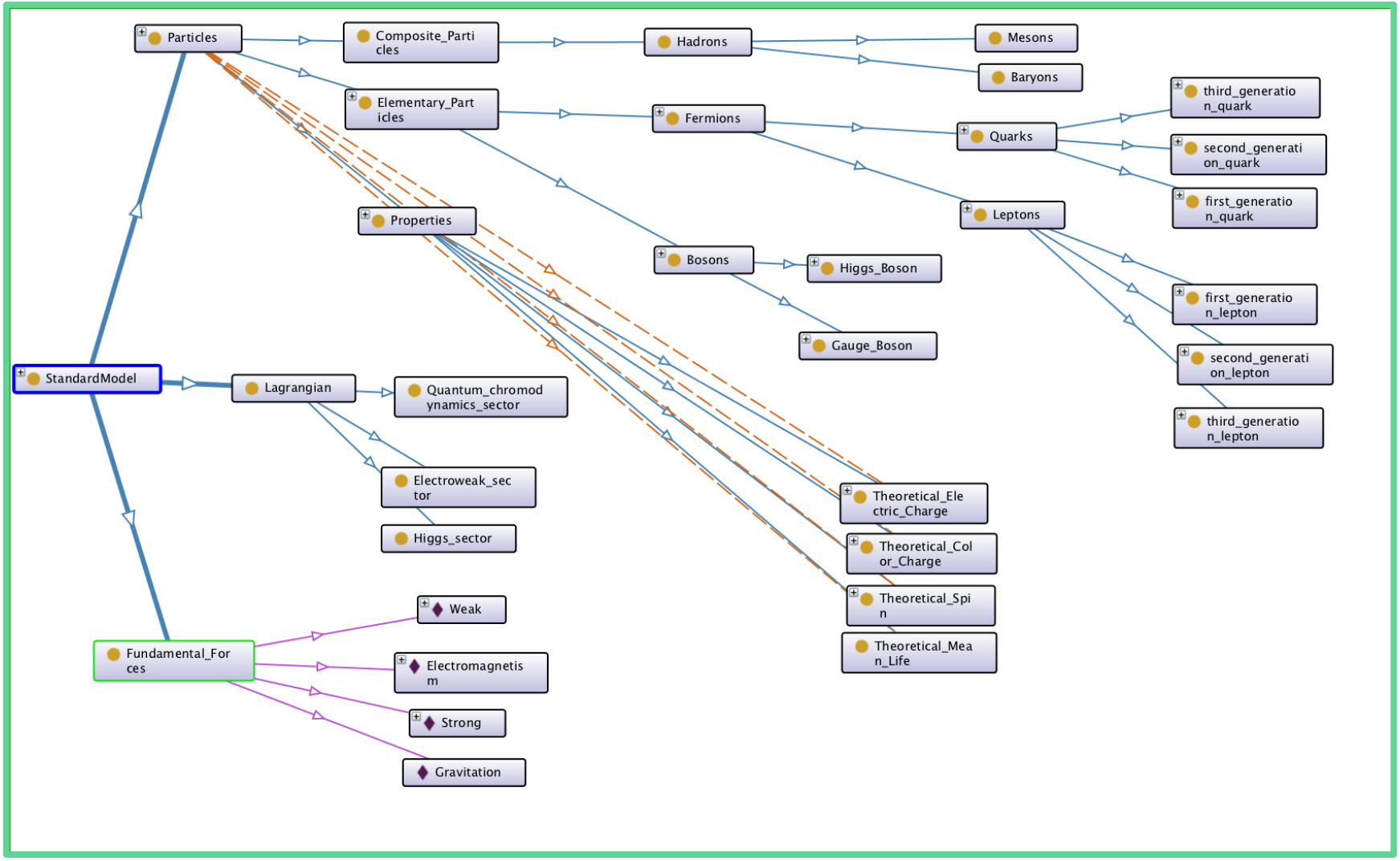
Documentation

Includes different types of documents required for preservation

- Discussion
- Internal Note
- Presentation
- Publication

CMS Open Data Ontology Structure: Individuals and Properties




Particles
Composite_Particles
Hadrons
Mesons
Baryons
Elementary_Particles
Fermions
Quarks
third_generation_quark
second_generation_quark
first_generation_quark
Leptons
first_generation_lepton
second_generation_lepton
third_generation_lepton
Properties
Bosons
Higgs_Boson
Gauge_Boson
StandardModel
Lagrangian
Quantum_chromodynamics_sector
Electroweak_sector
Higgs_sector
Theoretical_Electric_Charge
Theoretical_Color_Charge
Theoretical_Spin
Theoretical_Mean_Life
Fundamental_Forces
Weak
Electromagnetism
Strong
Gravitation

CMS Open Data Ontology Structure: SPARQL Query

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX onto: <http://www.semanticweb.org/guadalupecanasherrera/ontologies/2015/8/COD_Ontology#>
SELECT ?individuals
        WHERE { ?individuals rdf:type onto:Quarks }
```

individuals
down
charm
bottom
top
strange
up

CMS Open Data Ontology Structure: SPARQL Query

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX onto: <http://www.semanticweb.org/guadalupecanasherrera/ontologies/2015/8/COD_Ontology#>
SELECT *
WHERE {?s onto:experiences ?x}
```

s	x
down	Strong
strange	Strong
electron	Weak
tau	Weak
electron	Electromagnetism
bottom	Strong
muon	Weak
charm	Strong
top	Strong
gluon	Strong
muon	Electromagnetism
up	Strong
tau	Electromagnetism
tau_neutrino	Weak
electron_neutrino	Weak

The problems

Lack of previous complete examples

- HEP Complex system
- Blurred horizons for uses
- Difficult interaction with the Ontology without the proper Graphical Interface

The solutions?

Application to students

My proposal: apply the use of an ontology to Bachelor students starting in HEP with a proper Graphical Interface and study how useful they find the application

User’s Guide

Thank you for your attention

Any questions, suggestions, or improvements?

Who we are

Palmerina González Izquierdo

pgi25@alumnos.unican.es

https://github.com/Palmerina/CmsOpenData_IFCA

Guadalupe Cañas Herrera

gchi24@alumnos.unican.es

https://github.com/gcanasherrera/CmsOpenData_IFCA