



# **Implementación de un portal de datos (colisiones en CMS-LHC) para uso didáctico**

---

## **MEMORIA DE PRÁCTICAS**

INSTITUTO DE FÍSICA DE CANTABRIA  
Departamento de Física de Partículas experimental  
Av. de los Castros, 39005 Santander, Cantabria

Autor:  
**Palmerina González Izquierdo**

Directores:  
**Jesús Marco de Lucas**  
**Alicia Calderón Tazón**

Grado en Física

Santander, 18/08/2015 - 30/09/2015

## 1. ENTIDAD COLABORADORA:

Las prácticas han sido realizadas en el Instituto de Física de Cantabria (IFCA), uno de los cuatro institutos de investigación de la Universidad de Cantabria. Este centro es un referente para la investigación en Cantabria, colaborando en proyectos tan importantes como ha sido, por ejemplo, la búsqueda del bosón de Higgs en el CERN. El IFCA se dedica a investigar en diversas ramas de la física, como la astrofísica, la física de partículas, la fotónica, la meteorología, etc. En mi caso, fui asignada al departamento de Física de Partículas experimental.

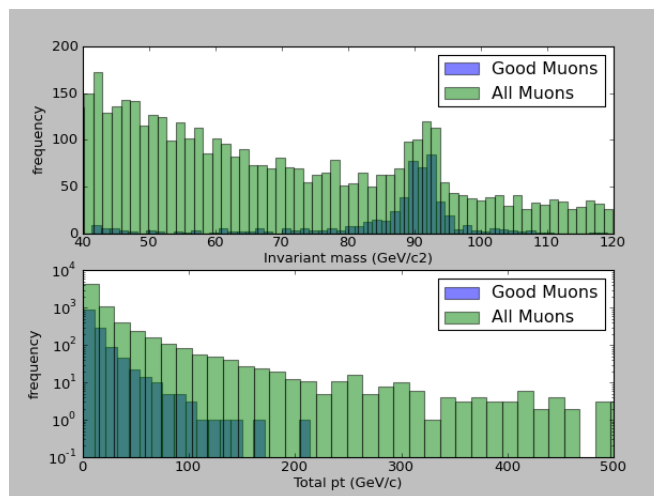
## 2. TAREAS DESEMPEÑADAS:

Mi trabajo ha consistido en la creación de un software de análisis de datos del CMS (Compact Muon Solenoid, detector de partículas del CERN) para fines didácticos. Estos datos corresponden a datos públicos de CMS del año 2010.

Mi función ha sido la de actualizar el software de la ya existente máquina virtual CMS *OPEN DATA PORTAL @IFCA*, destinado a la enseñanza de física de partículas básica (la página web donde se accede a esta máquina virtual es: <https://cmsopendata.ifca.es/>).

En principio, lo que se buscaba era desarrollar un programa con código más simple al anterior para que los estudiantes pudiesen entenderlo. Con este mismo objetivo en mente, el lenguaje de programación escogido fue python para poder implementar el código en una hoja de *ipython notebook*. Este entorno de desarrollo computacional permite combinar la ejecución del código, insertar texto explicativo, crear los gráficos del programa e insertar imágenes. De esta forma, los estudiantes tendrían una interfaz en la que podrían seguir los pasos indicados y modificar algunas variables para obtener las gráficas deseadas y ver así los datos de las colisiones de partículas. Sin esta herramienta, los estudiantes tendrían que conectarse en remoto a la máquina virtual e ir modificando el código directamente con comandos de la terminal, lo cual es muy engorroso si es la primera vez que se usa.

Para ilustrar mejor en qué consistió mi trabajo, en la Fig. 1 muestro un ejemplo de las gráficas que deberían obtener los estudiantes al modificar correctamente los valores indicados. En este caso, los usuarios del programa deberían variar una serie de valores numéricos, que corresponden a los cortes de selección de los muones que vienen del decaimiento del bosón Z (por ejemplo el momento transversal mínimo que deben tener, la distancia máxima desde donde se originan al punto de colisión de las partículas, etc.). Al poner unos valores razonables, obtendrían el pico que se ve en la Fig. 1.



**Figura 1:** Histogramas obtenidos con mi software CMS\_OpenData\_DidacticAnalyzer que muestra la masa invariante y el momento transversal de los pares de muones sin aplicar los cortes de selección (verde) y tras aplicarlos (azul).

### 3. PROBLEMAS Y SU RESOLUCIÓN:

Habiendo ya desarrollado el código y comprobado su correcto funcionamiento, se organizó una reunión con el resto del grupo para explicar mi trabajo y la forma de implementarlo en *ipython notebook*.

En la susodicha reunión nos percatamos de un problema bastante grave: el *ipython notebook* necesita tener instalada una versión de python superior a la 2.7, y la disponible en la máquina virtual es la 2.6. Asimismo, no era posible actualizar la versión ya que los datos públicos del CMS son del año 2010, los cuales necesitan ser leídos utilizando un software específico llamado CMSSW. La versión de CMSSW usada en el 2010 necesita la versión de python 2.6, por lo que si la actualizábamos en la máquina virtual, mi programa no funcionaría.

Este problema se abordó por dos frentes:

En primer lugar, procedí a cambiar el formato de los datos, pasando de PAT-Tuplas a N-Tuplas, las cuales no necesitan del CMSSW software para ser leídas. Además, fue necesario crear otro programa que analizara estos datos, similar al que ya había hecho.

Sin embargo, debido al poco tiempo del que disponía y la elección de python como lenguaje para convertir los datos (este proceso suele implementarse en C++), el programa, aunque la estructura esté escrita, está inacabado. Por esto, también ha sido mi responsabilidad proporcionar mi código (se encuentra en el repositorio: [https://github.com/Palmerina/CMS\\_OpenData\\_DidacticAnalyzer/tree/master/CmsOpenData\\_Analyzer](https://github.com/Palmerina/CMS_OpenData_DidacticAnalyzer/tree/master/CmsOpenData_Analyzer)), explicar todo el procedimiento seguido y los objetivos que se quieren conseguir a otra estudiante en prácticas, que continuará con el trabajo.

En segundo lugar, otra estudiante en prácticas que se encarga de crear una ontología del modelo estándar siguiendo la estructura de mi programa y yo decidimos crear una guía

de usuario documentando ambas el programa y la ontología (se puede ver una muestra de la guía en la Fig. 2). En ella, explicamos la conexión a la máquina virtual en remoto a través de ssh, cómo acceder a los códigos para modificarlos en la terminal a través del editor de texto *vim*, comandos útiles de la terminal, cómo ejecutar los ejercicios propuestos, el funcionamiento detallado del código, etc. También se incluyeron los conceptos físicos necesarios para utilizar el programa aunque, una vez más, debido a la limitación en el tiempo, esta sección quedó incompleta. No obstante, este documento se encuentra compartido en <https://www.overleaf.com> por las tres estudiantes involucradas en el proyecto para continuar añadiendo información.



Figura 2: Portada y tercer capítulo de la guía de usuario.

#### 4. VALORACIÓN Y APORTACIÓN AL APRENDIZAJE:

Considero que estas prácticas han contribuido en gran medida a mi aprendizaje, tanto en conocimientos como en habilidades necesarias para trabajar en una empresa u organismo público en un futuro. De la misma forma, los conocimientos adquiridos en la carrera sobre física de partículas y programación me han resultado indispensables a la hora de abordar el trabajo.

No sólo he profundizado en física de partículas y he aprendido un lenguaje de programación completamente nuevo (*python*), para lo cual seguí un curso de edX (*Introduction to Computer Science and Programming Using Python*) impartido por el MIT; también he desarrollado gran cantidad de nuevas habilidades útiles de cara al futuro, como el uso de la terminal, *vim*, una máquina virtual, el uso de un repositorio, etc.

Por último, me ha enriquecido mucho la experiencia de ver cómo se trabaja en un

centro de investigación, ya que es una de las posibles salidas profesionales que más me atre.

##### 5. EVALUACIÓN DE LAS PRÁCTICAS Y SUGERENCIAS DE MEJORA:

Como ya he expresado anteriormente, estas prácticas han resultado muy positivas para mí, ya que han contribuido enormemente a mi aprendizaje y me han introducido en el mundo laboral, ya que hasta ahora sólo me había dedicado a mis estudios. De esta forma, las prácticas me han permitido contemplar en primera persona en qué consiste trabajar en un centro de investigación, facilitándome la elección de qué voy a proponerme ser cuando termine el grado.

El único inconveniente que he visto ha sido la limitación de tiempo, ya que me gustaría haber terminado todo lo que tenía pensado hacer. Por otra parte, también cabe reseñar que el hecho de tener a mis dos tutores (el de la empresa y el de la universidad) en el IFCA supervisando mi trabajo, conllevó a malentendidos por falta de comunicación entre las partes. Sin embargo, no considero que esto entorpeciera el trabajo de forma notable.