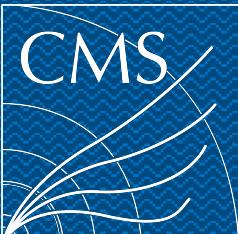


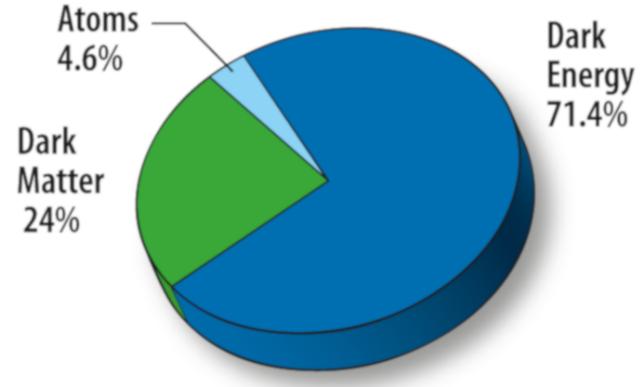


# On-Demand Distributed Workflows for Physics Analysis at the CMS Experiment

Diyaselis M. Delgado López

Department of Physics  
UPRM CMS Research Group



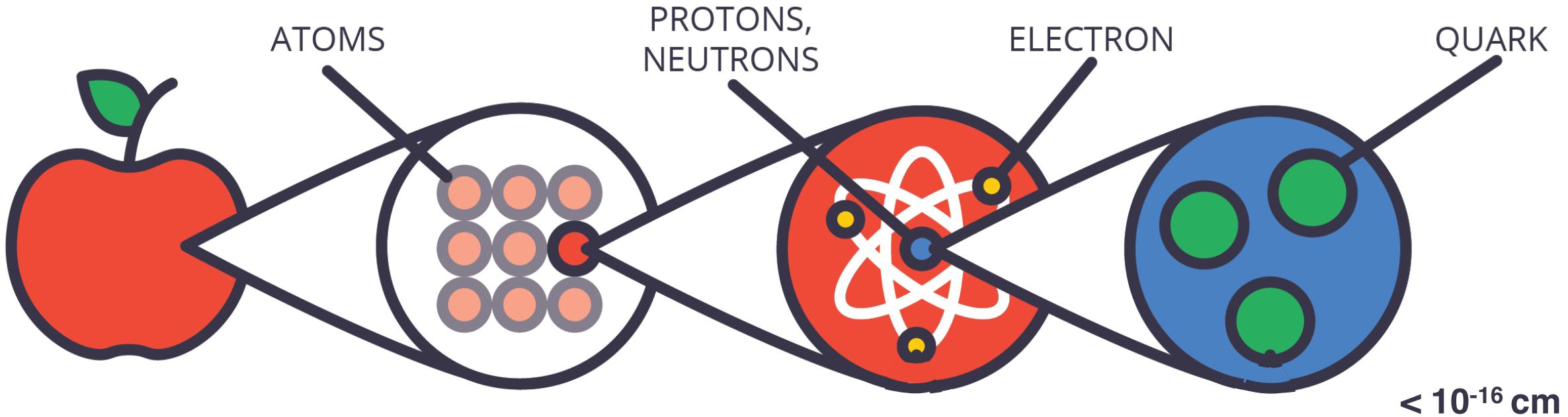


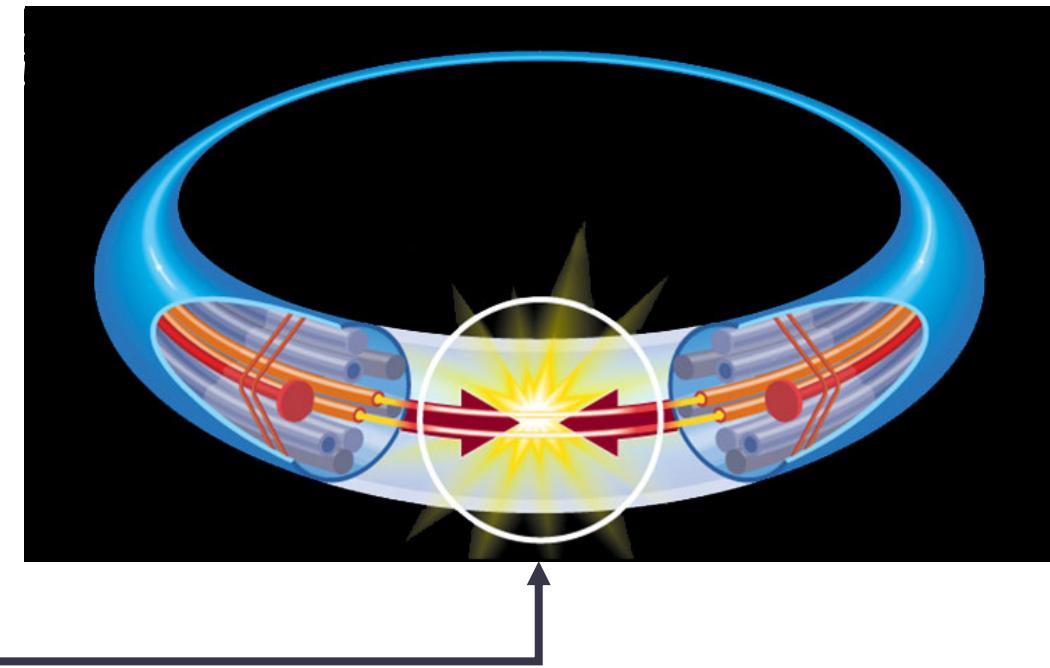
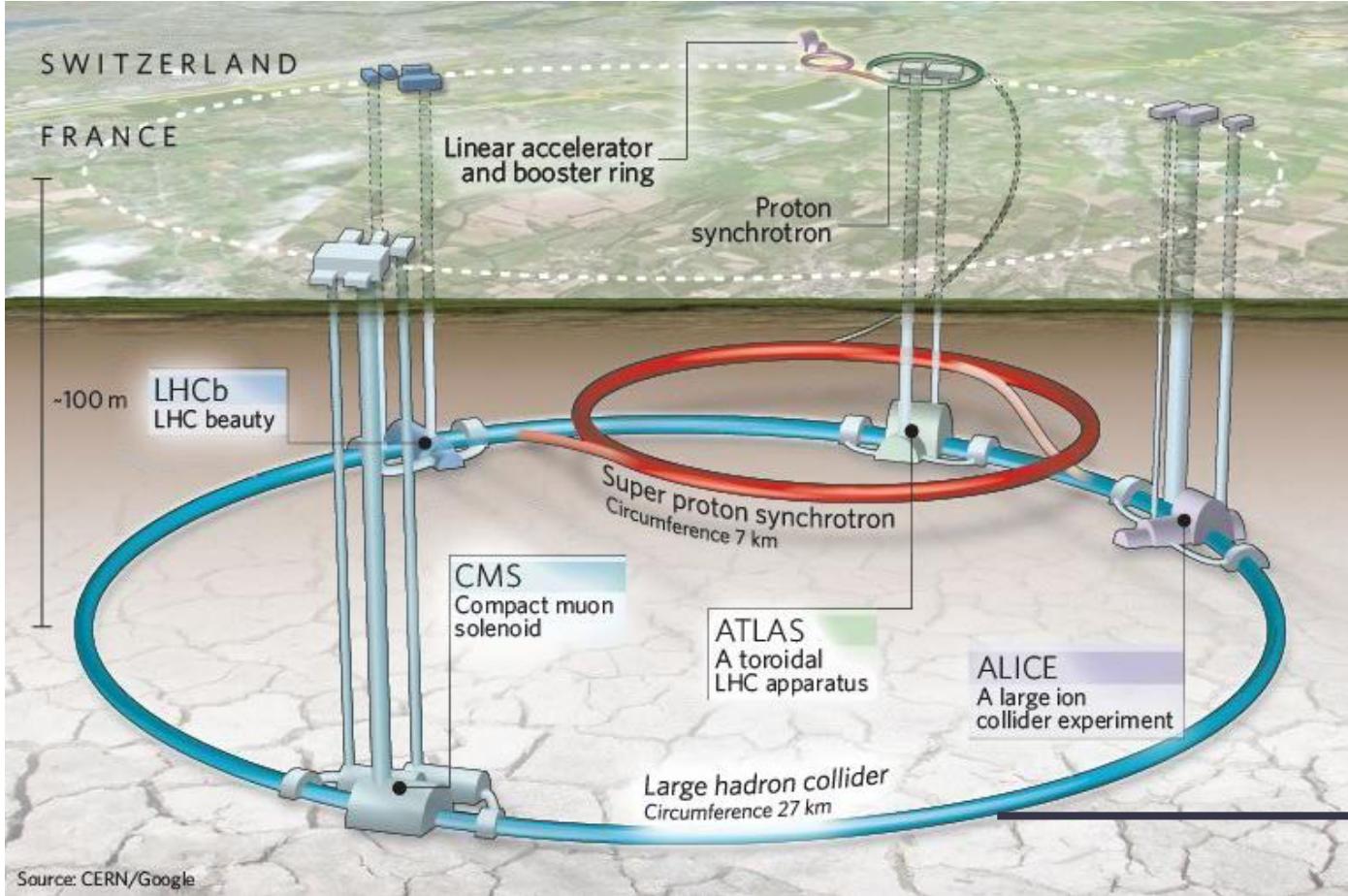
# Let's have a look at the Universe?

**MATTER**

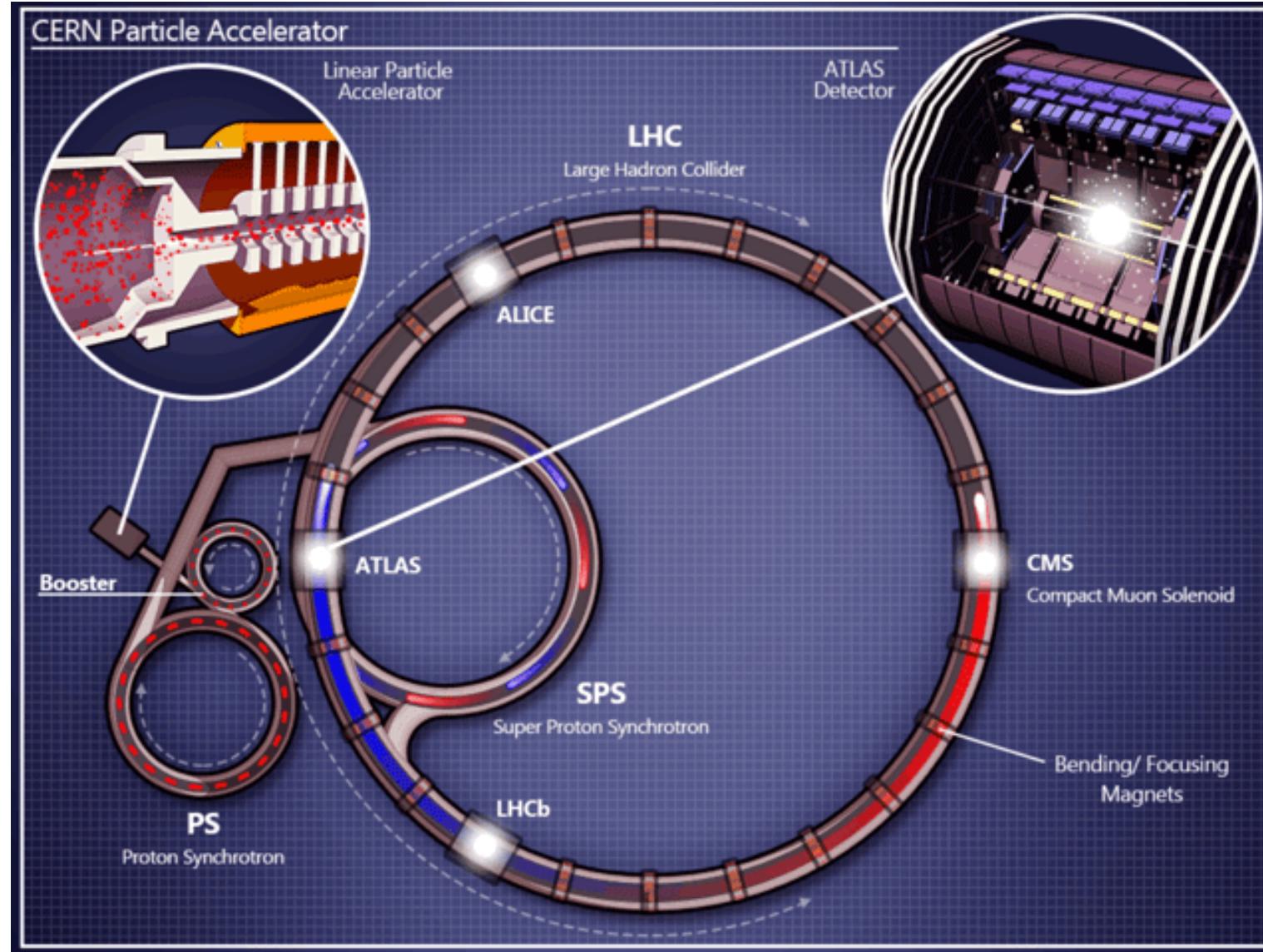
**ATOM**

**NUCLEUS**





**Large Hadron Collider** is the main 27-kilometer particle accelerator ring at **CERN**.



Accelerators boost **beams of particles** to high energies and are made to collide with each other.

# Compact Muon Solenoid

## CMS DETECTOR



Total weight : 14,000 tonnes  
Overall diameter : 15.0 m  
Overall length : 28.7 m  
Magnetic field : 3.8 T

### STEEL RETURN YOKE

12,500 tonnes

### SILICON TRACKERS

Pixel ( $100 \times 150 \mu\text{m}$ )  $\sim 16\text{m}^2 \sim 66\text{M}$  channels  
Microstrips ( $80 \times 180 \mu\text{m}$ )  $\sim 200\text{m}^2 \sim 9.6\text{M}$  channels

### SUPERCONDUCTING SOLENOID

Niobium titanium coil carrying  $\sim 18,000\text{A}$

### MUON CHAMBERS

Barrel: 250 Drift Tube, 480 Resistive Plate Chambers  
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

### PRESHOWER

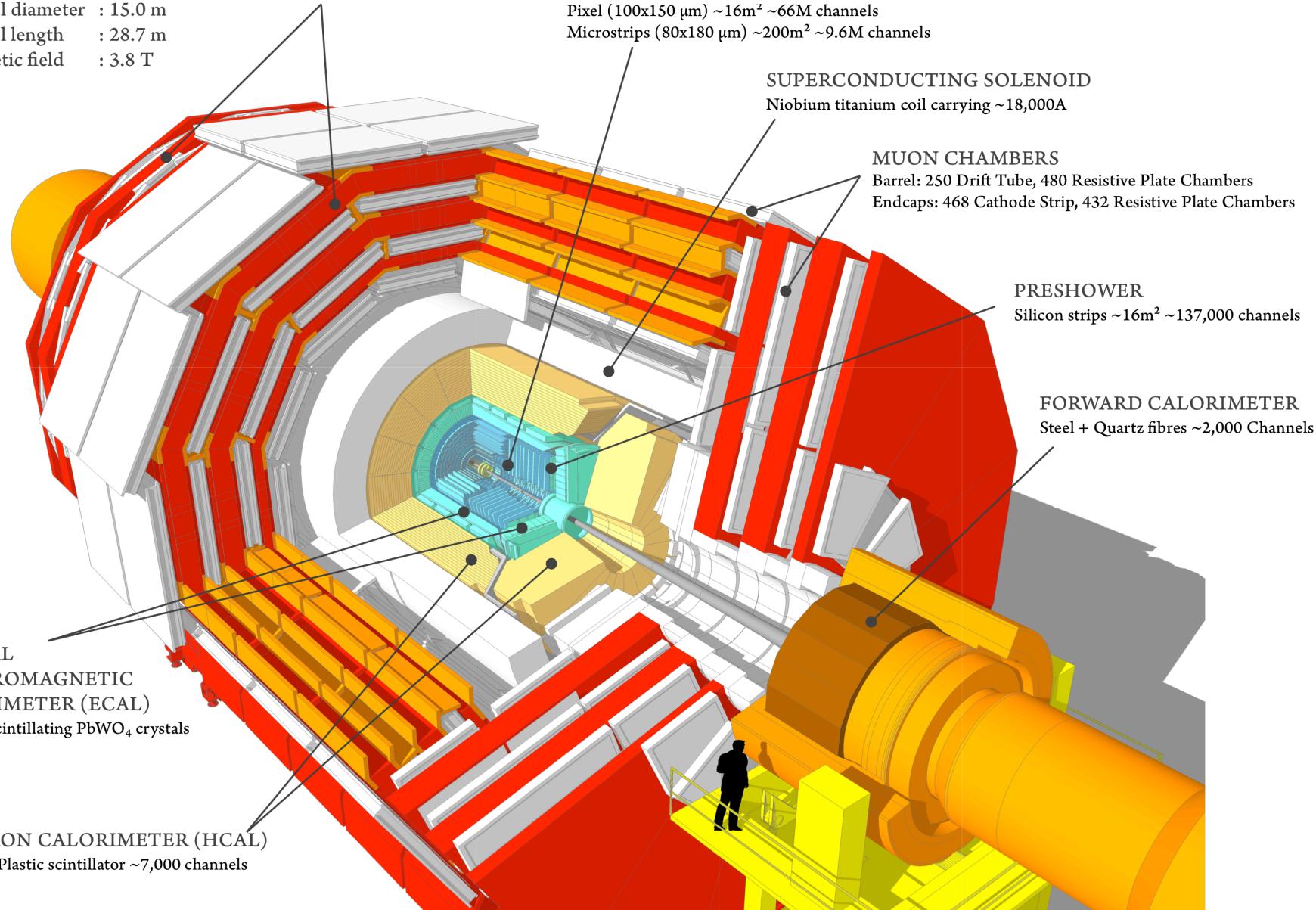
Silicon strips  $\sim 16\text{m}^2 \sim 137,000$  channels

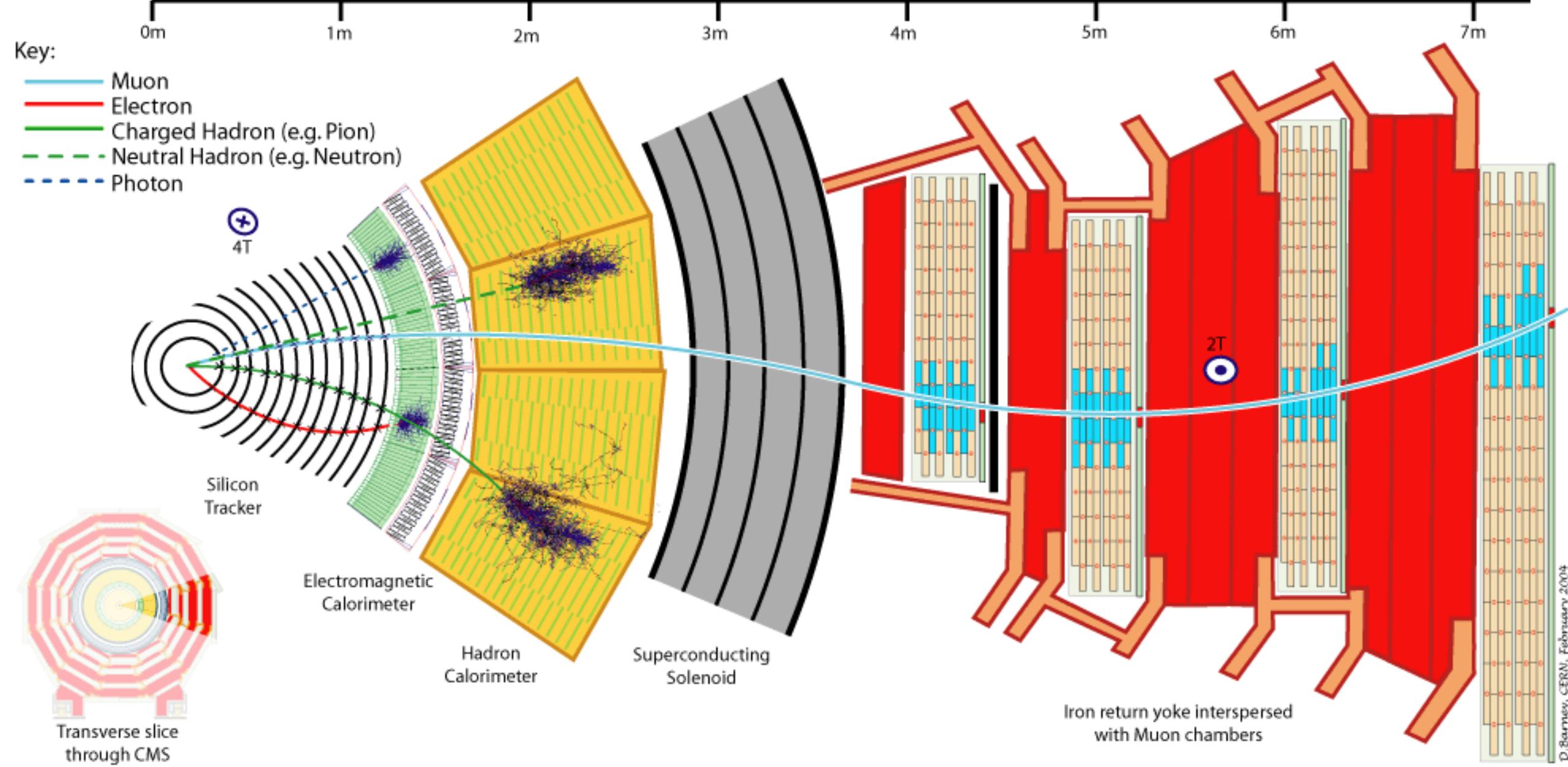
### FORWARD CALORIMETER

Steel + Quartz fibres  $\sim 2,000$  Channels

CRYSTAL  
ELECTROMAGNETIC  
CALORIMETER (ECAL)  
 $\sim 76,000$  scintillating  $\text{PbWO}_4$  crystals

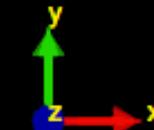
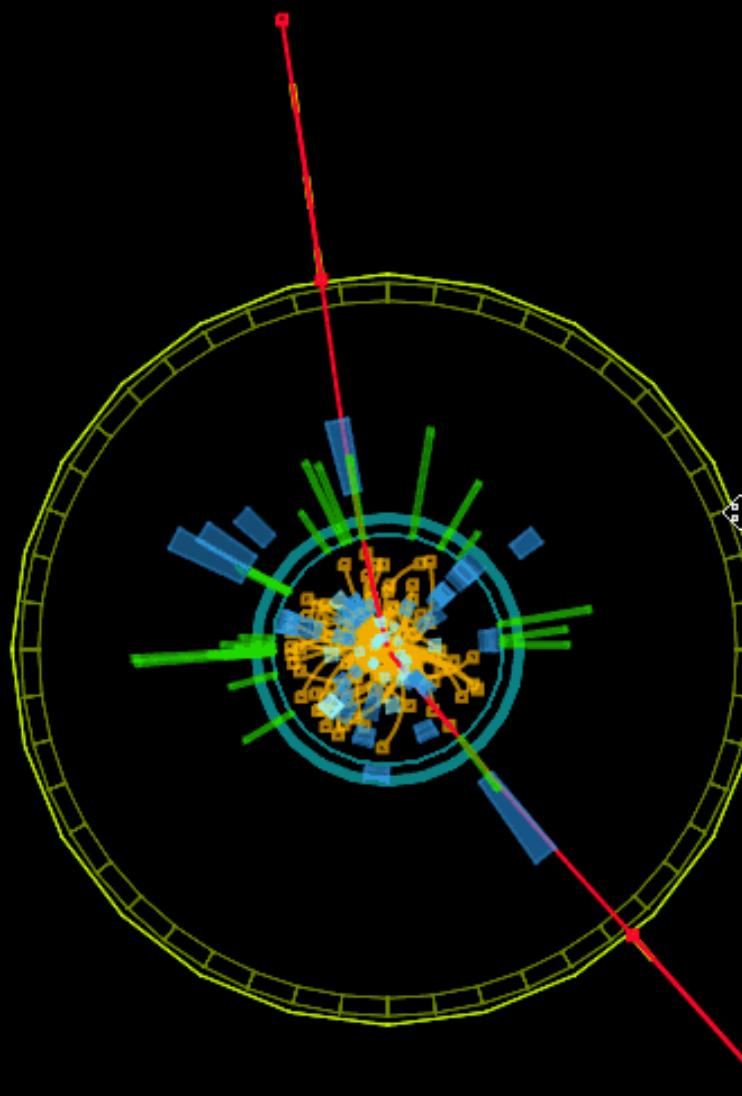
HADRON CALORIMETER (HCAL)  
Brass + Plastic scintillator  $\sim 7,000$  channels







| Detector Model                  |                                       |
|---------------------------------|---------------------------------------|
| Tracker Barrels                 | <input type="checkbox"/>              |
| Tracker Endcaps                 | <input type="checkbox"/>              |
| ECAL Barrel                     | <input checked="" type="checkbox"/>   |
| ECAL Endcaps                    | <input type="checkbox"/>              |
| ECAL Preshower                  | <input type="checkbox"/>              |
| HCAL Barrel                     | <input type="checkbox"/>              |
| HCAL Endcaps                    | <input type="checkbox"/>              |
| HCAL Outer                      | <input checked="" type="checkbox"/>   |
| HCAL Forward                    | <input type="checkbox"/>              |
| Drift Tubes (muon)              | <input type="checkbox"/>              |
| Cathode Strip Chambers (muon)   | <input type="checkbox"/>              |
| Resistive Plate Chambers (muon) | <input type="checkbox"/>              |
| Tracking                        |                                       |
| Tracks (reco.)                  | <input checked="" type="checkbox"/>   |
| Clusters (Si Pixels)            | <input type="checkbox"/>              |
| Clusters (Si Strips)            | <input type="checkbox"/>              |
| Rec. Hits (Tracking)            | <input type="checkbox"/>              |
| ECAL                            |                                       |
| Barrel Rec. Hits                | <input checked="" type="checkbox"/> ▶ |
| Endcap Rec. Hits                | <input type="checkbox"/> ▶            |
| Preshower Rec. Hits             | <input type="checkbox"/> ▶            |
| HCAL                            |                                       |
| Barrel Rec. Hits                | <input checked="" type="checkbox"/> ▶ |
| Endcap Rec. Hits                | <input checked="" type="checkbox"/> ▶ |
| Forward Rec. Hits               | <input checked="" type="checkbox"/> ▶ |
| Outer Rec. Hits                 | <input type="checkbox"/> ▶            |



# DPOA

## Data Preservation and Open Access



# reana

- Reusable and reproducible research data analysis platform



**CERN**  
ANALYSIS PRESERVATION

- Addresses the need for the long-term preservation of the data analysis process

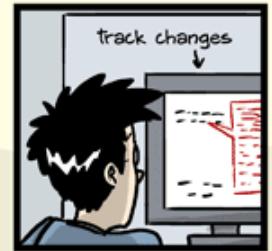
CASE STUDIES IN REPRESENTATION OF CMS

# "FINAL".doc



FINAL\_rev.6.COMMENTS.doc

FINAL\_rev.8.comments55.  
CORRECTIONS.doc



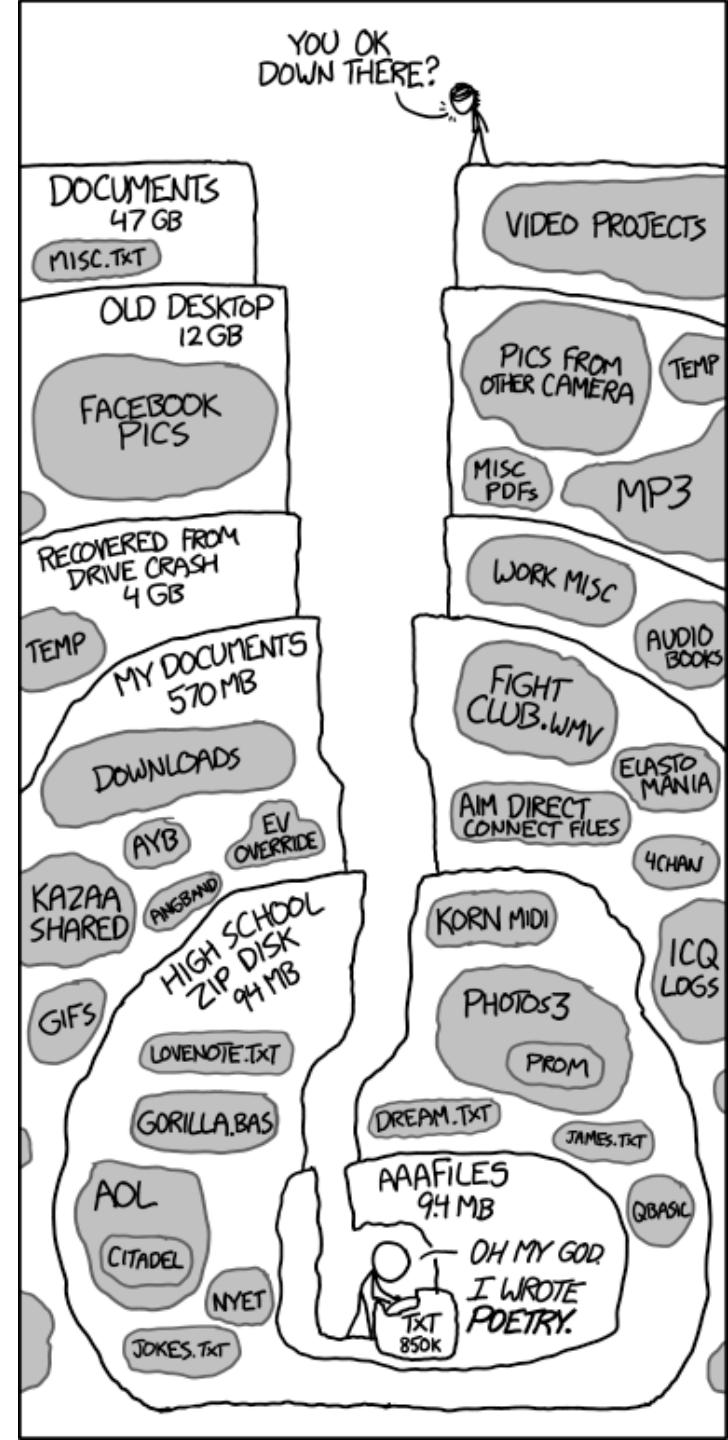
FINAL\_rev.18.comments7.  
corrections9.MORE.30.doc

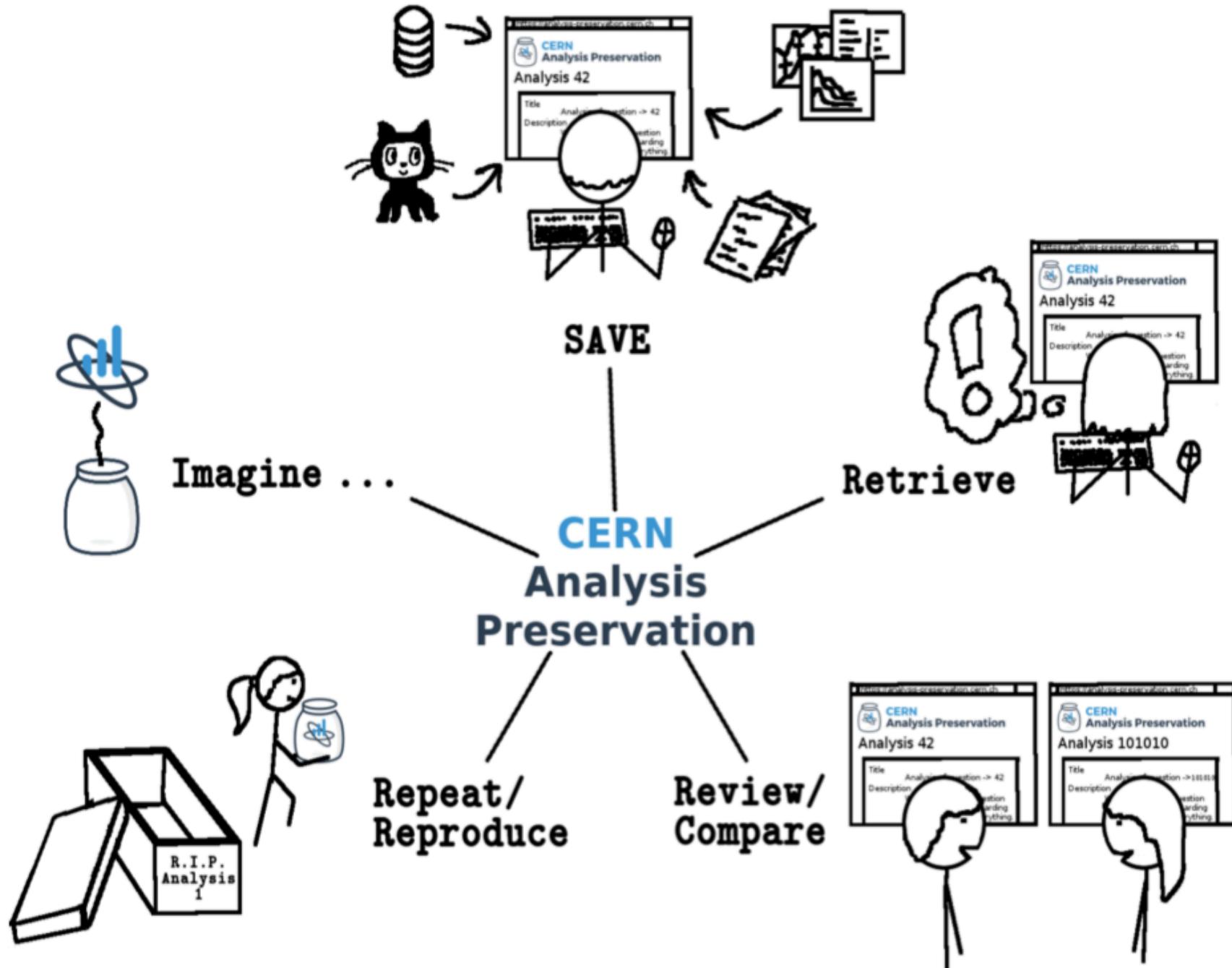
FINAL\_rev.22.comments49.  
corrections.10.#@\$%WHYDID  
ICOMETOGRAD SCHOOL?????.doc

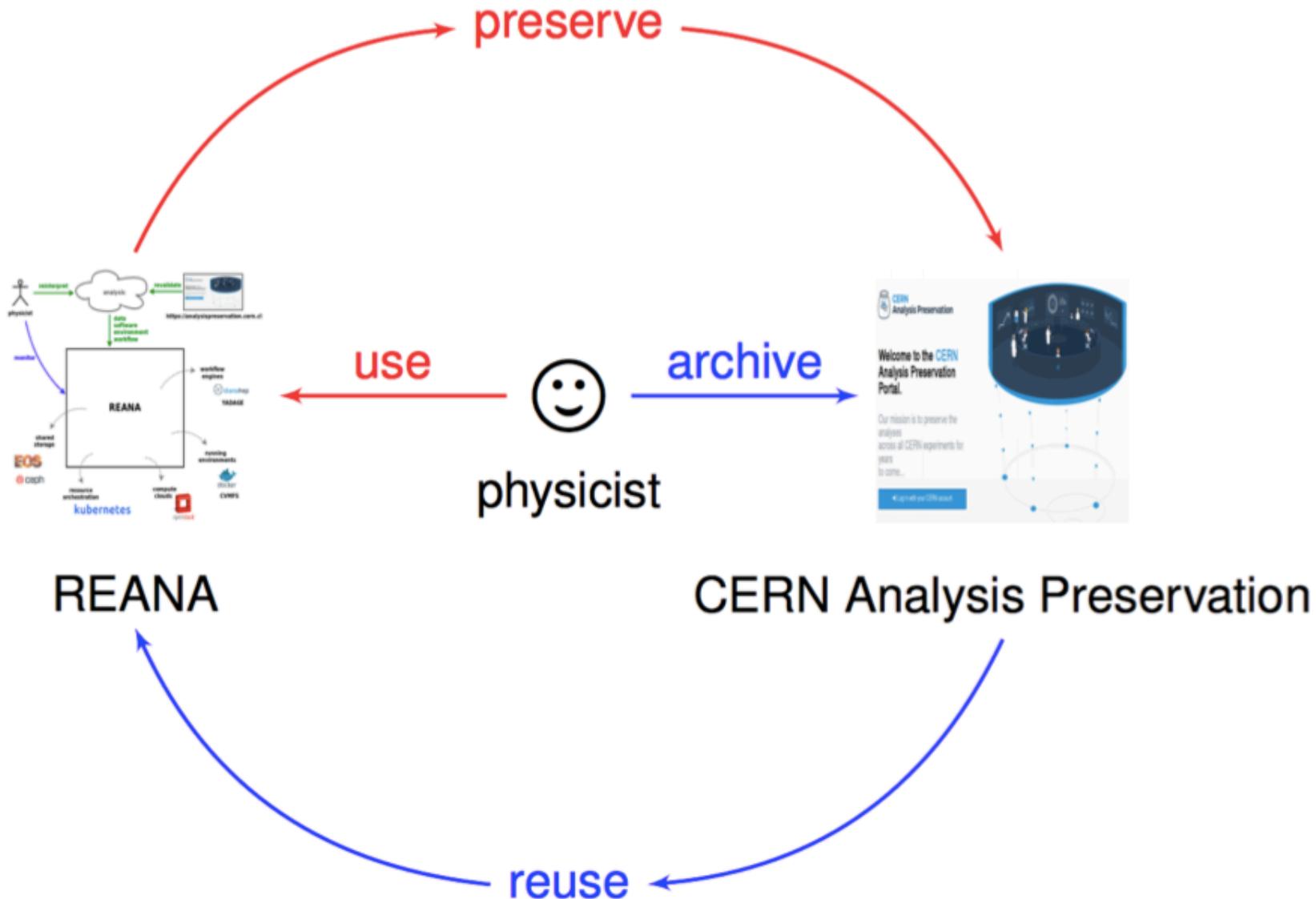
**Imagine ...**  
Ten years from now

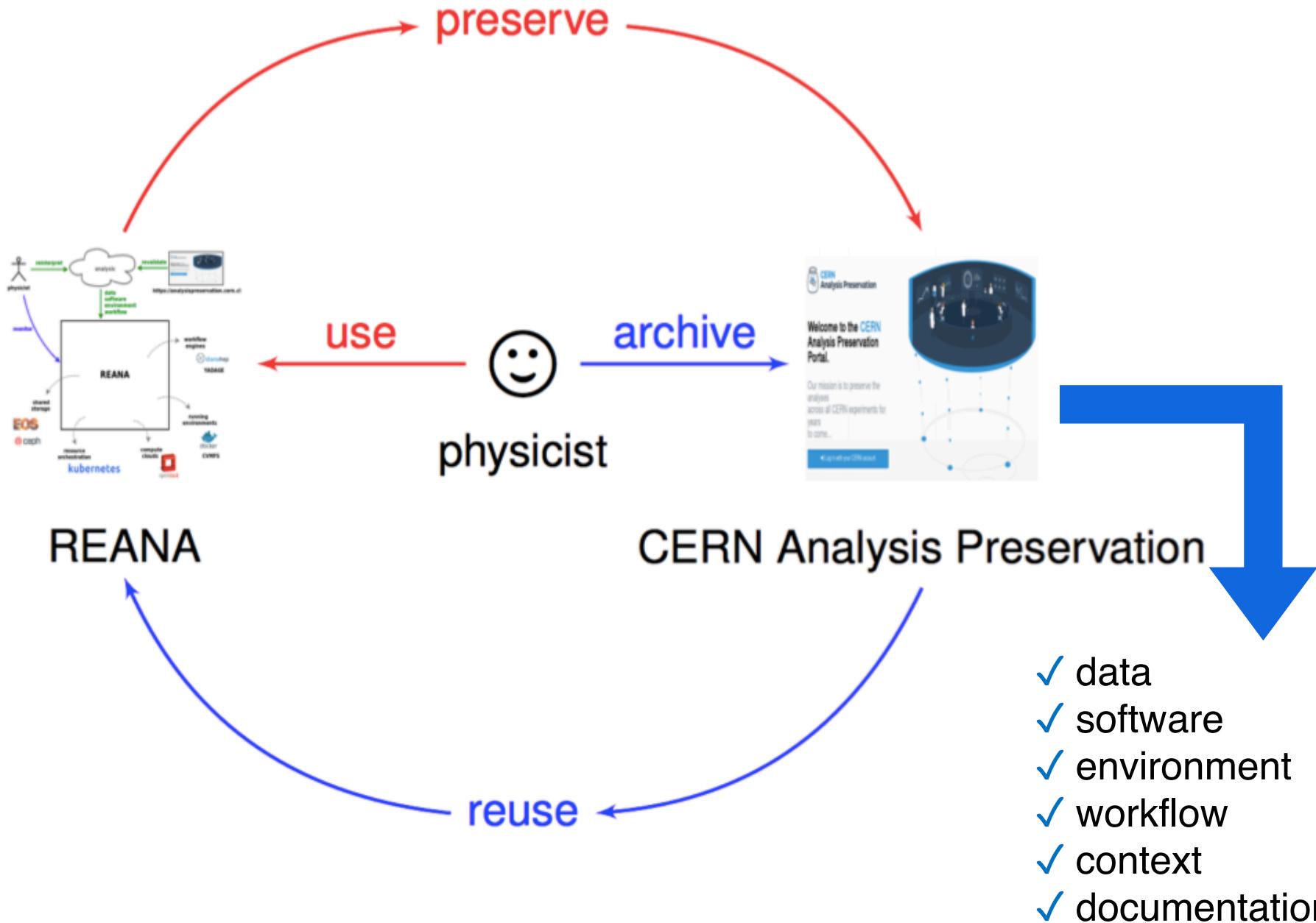
Would you rather  
have your analysis  
ready to be reused  
whenever you want  
or not?

YOU OK  
DOWN THERE?

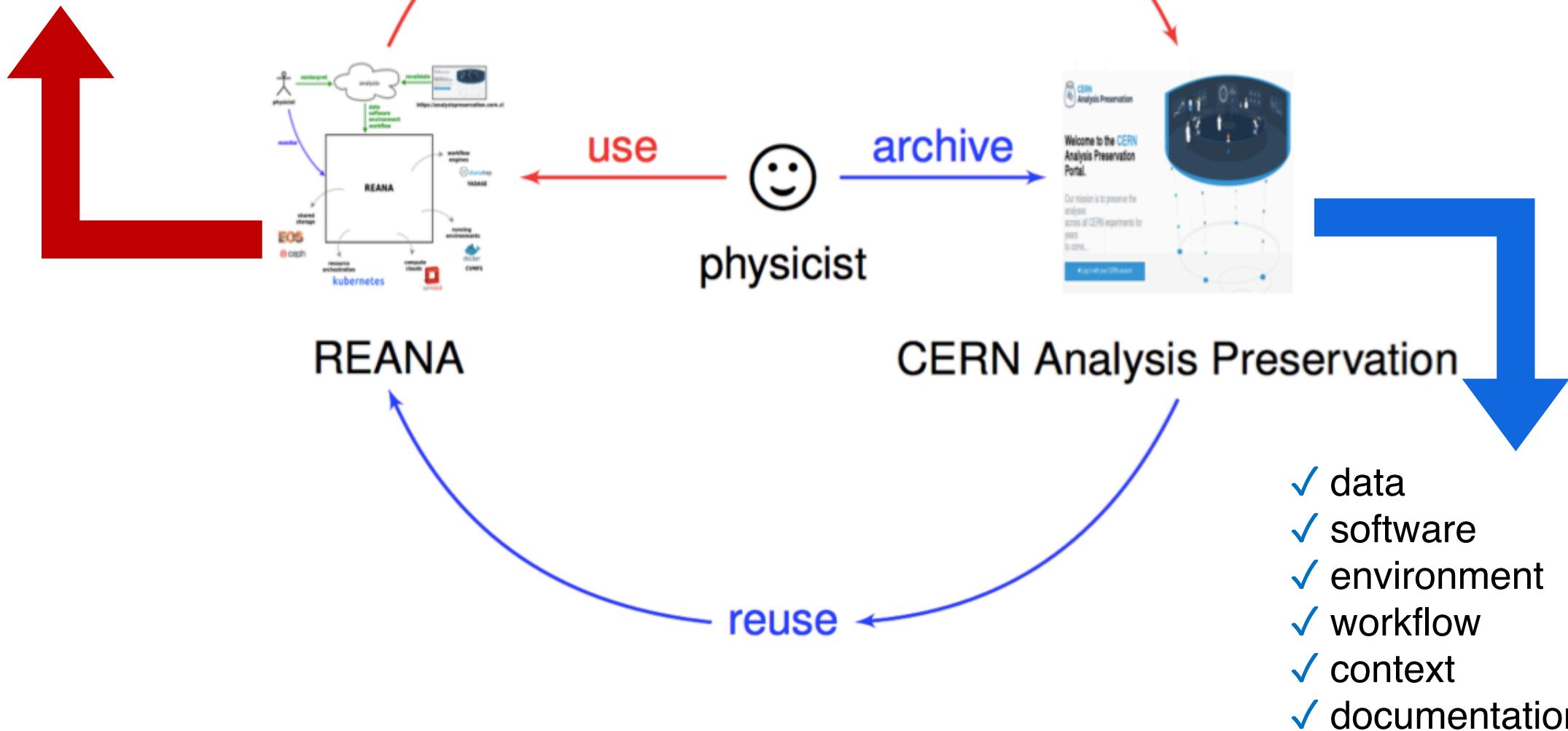








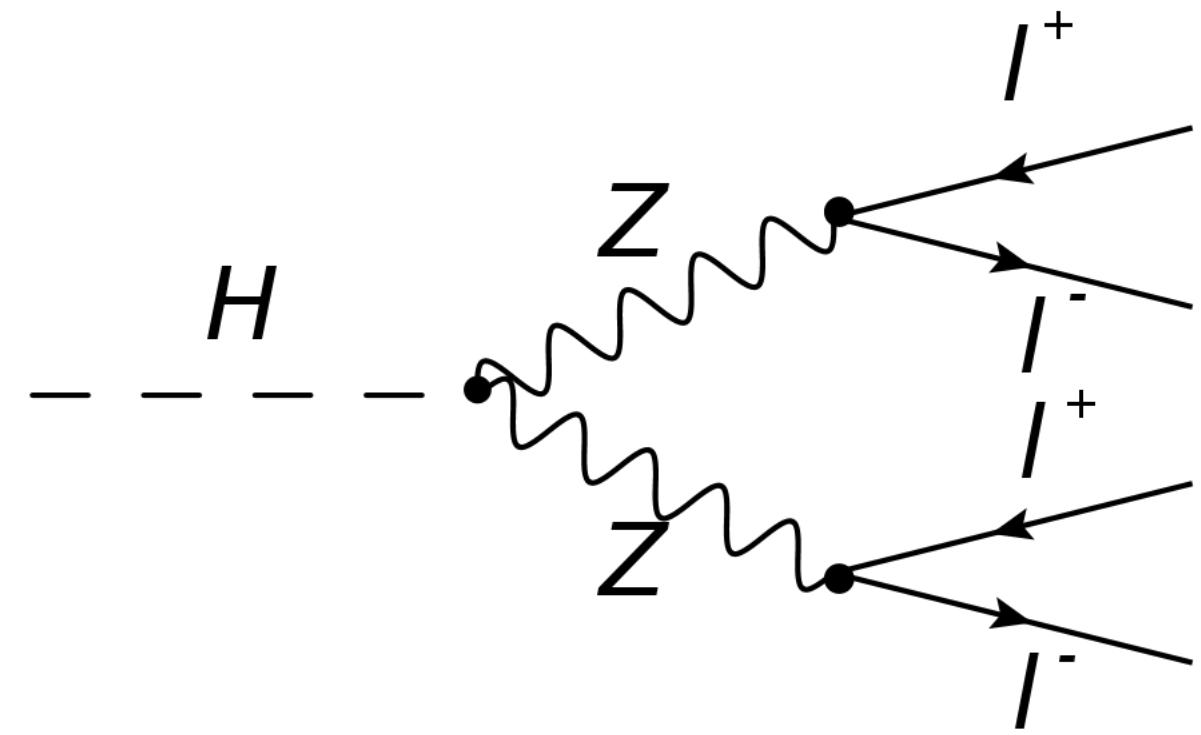
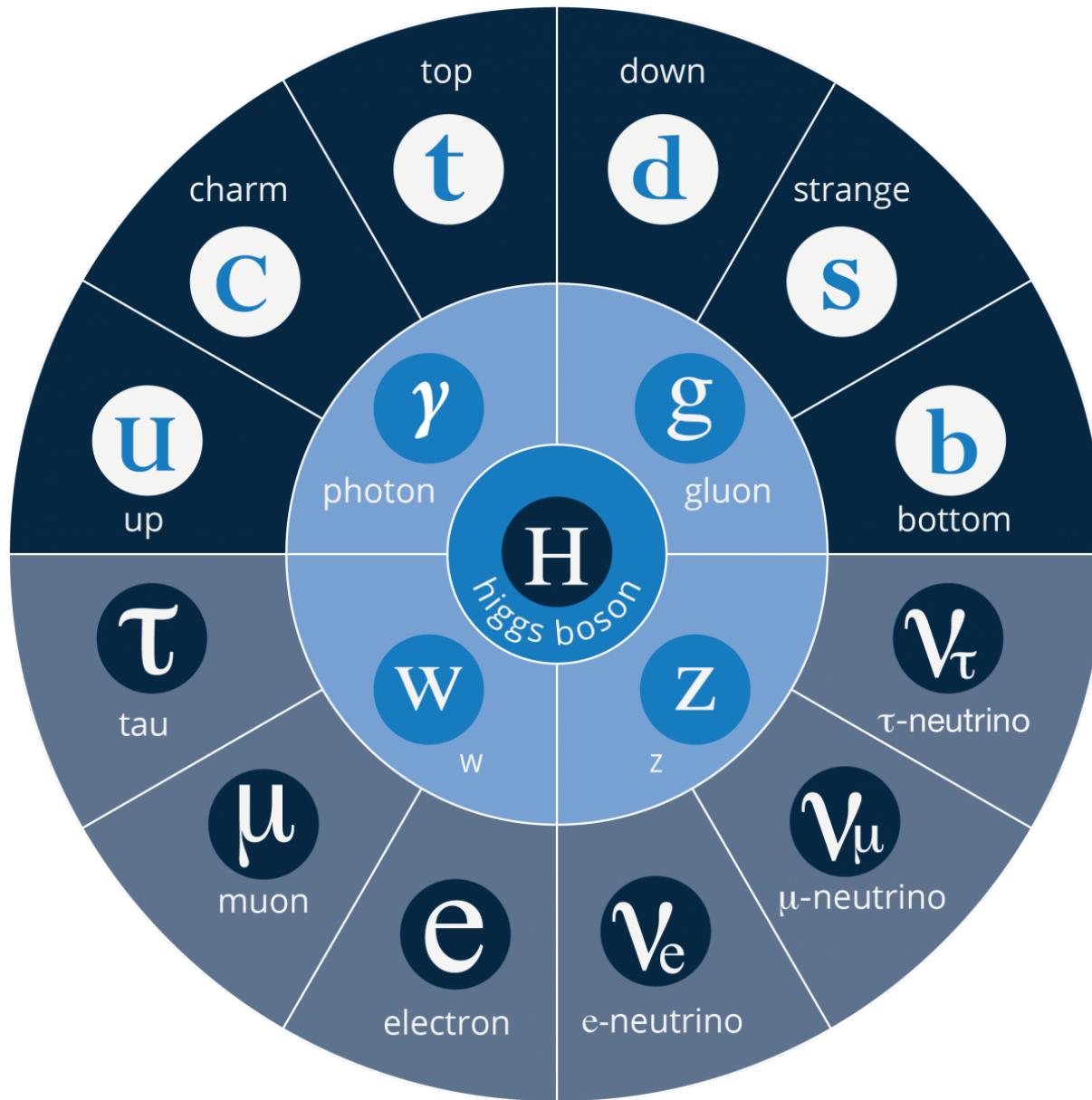
- ✓ computing clouds
- ✓ running environments
- ✓ resource orchestration
- ✓ workflow engines
- ✓ shared storage systems





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

# What does Higgs-to-four-leptons mean?



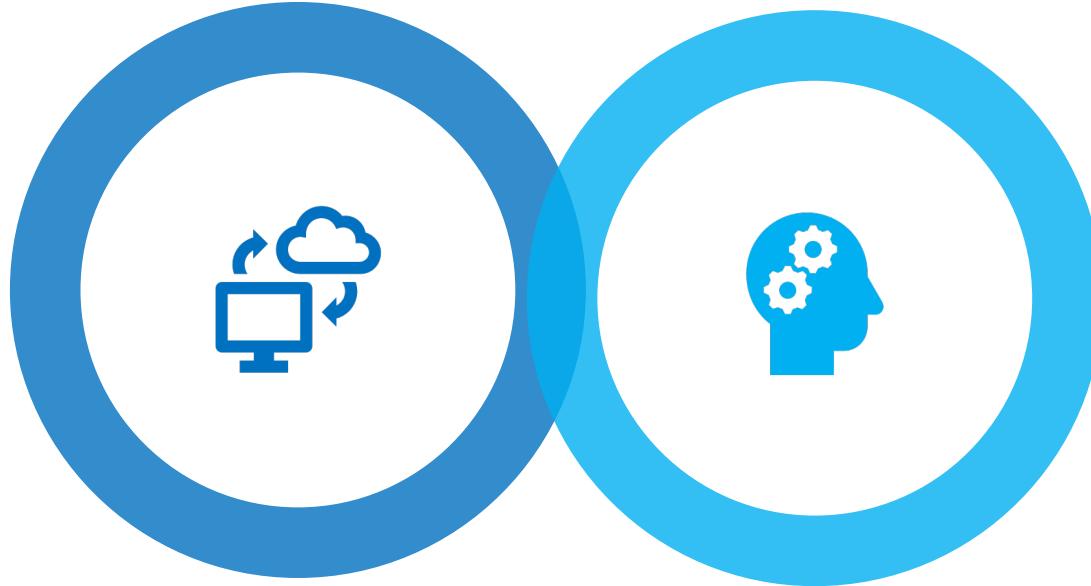
# Preservation and Reproducibility of the Analysis



**CMS Open Data**

Inputs and  
software  
environment

# Preservation and Reproducibility of the Analysis



## CMS Open Data

Inputs and  
software  
environment

## CAP

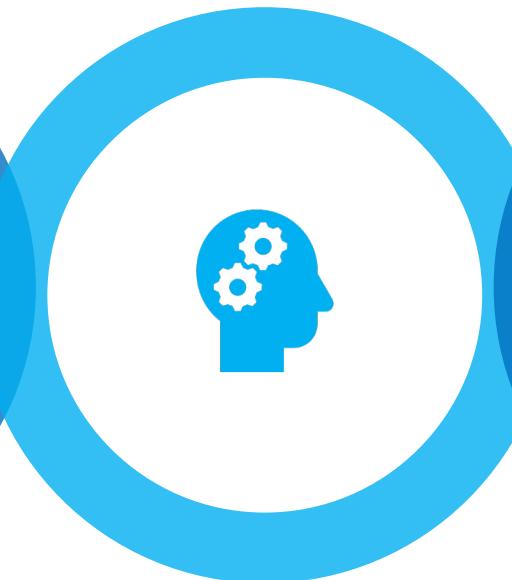
Analysis  
structure and  
metadata

# Preservation and Reproducibility of the Analysis



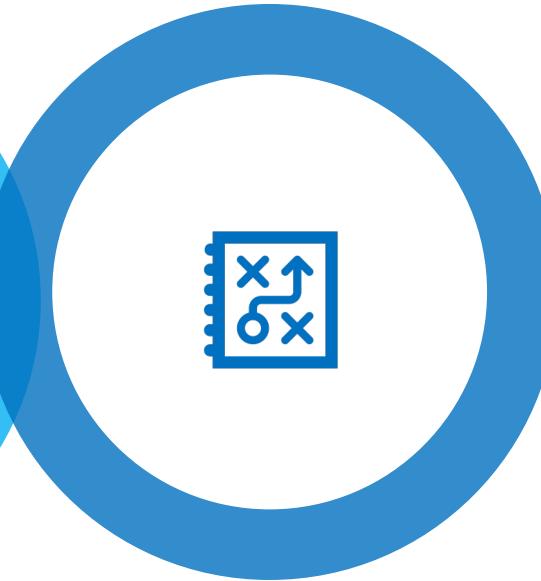
## CMS Open Data

Inputs and  
software  
environment



## CAP

Analysis  
structure and  
metadata



## ReANA

Analysis workflow,  
commands, and  
output

# Structure the Analysis

## 1 Input data

What is your input data?

- input files
- input parameters

## 2 Analysis code

Which code analyses it?

- software frameworks
- user code

## 3 Compute environment

What is your environment?

- operating system
- database calls

## 4 Analysis workflow

Which steps did you take?

- single command
- complex workflows

# Structure the Analysis

## 1 Input data

2011 – 2012  
RAW data and  
MC simulations

## 2 Analysis code

Which code analyses it?

- software frameworks
- user code

## 3 Compute environment

What is your environment?

- operating system
- database calls

## 4 Analysis workflow

Which steps did you take?

- single command
- complex workflows

# Structure the Analysis

## 1 Input data

2011 – 2012  
RAW data and  
MC simulations

## 2 Analysis code

Plot background  
and processed data

## 3 Compute environment

What is your environment?

- operating system
- database calls

## 4 Analysis workflow

Which steps did you take?

- single command
- complex workflows

# Structure the Analysis

## 1 Input data

2011 – 2012  
RAW data and  
MC simulations

## 2 Analysis code

Plot background  
and processed data

## 3 Compute environment

CMSSW image  
from Docker  
containers

## 4 Analysis workflow

Which steps did you take?

- single command
- complex workflows

# Structure the Analysis

## 1 Input data

2011 – 2012  
RAW data and  
MC simulations

## 2 Analysis code

Plot background  
and processed data

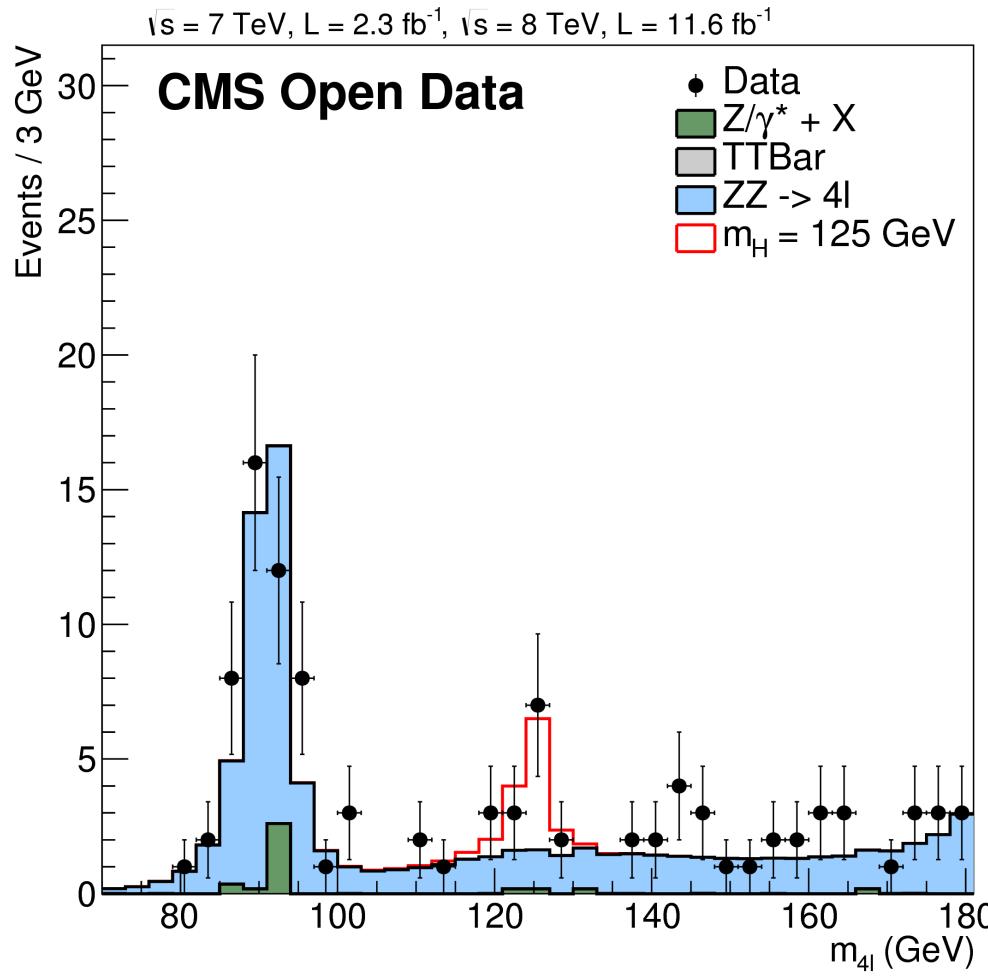
## 3 Compute environment

CMSSW image  
from Docker  
containers

## 4 Analysis workflow

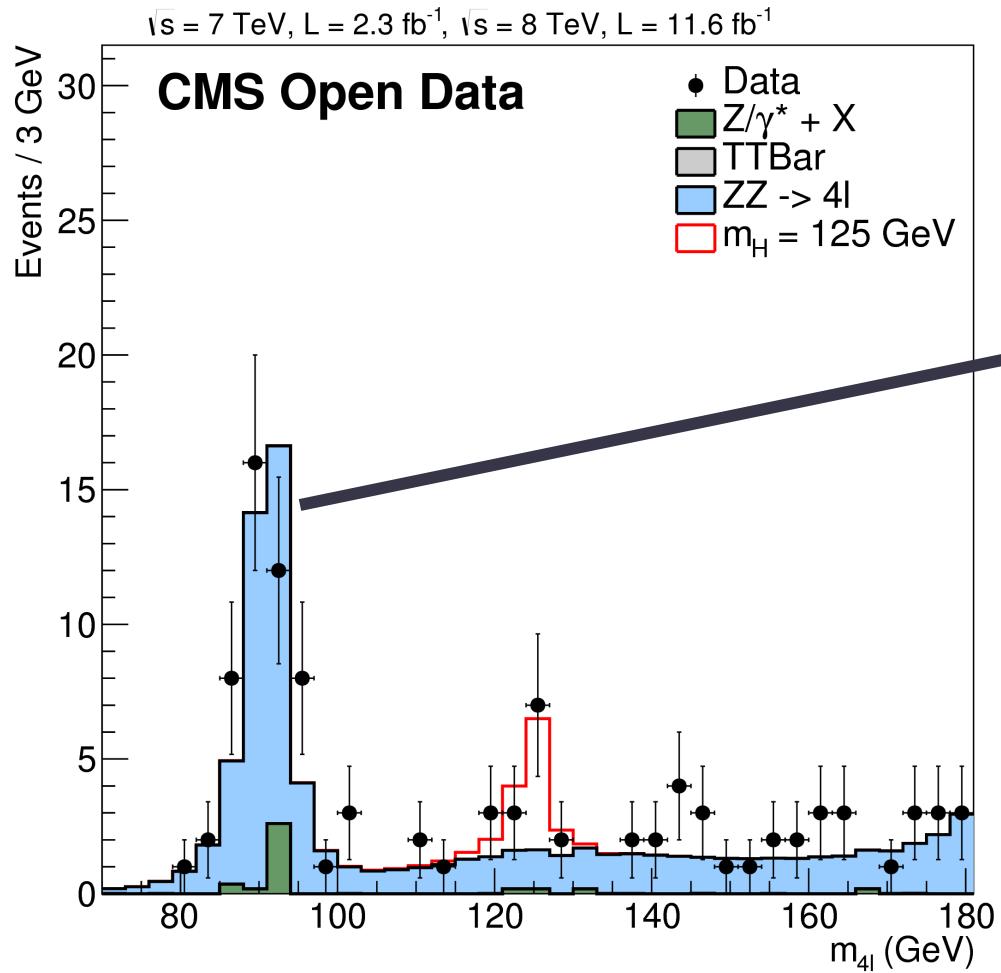
CWL (Common  
Workflow Language)

# Higgs discovery: simplified reimplementation



This data set is too small to say anything for certain, but it isn't too far off from actual analysis results.

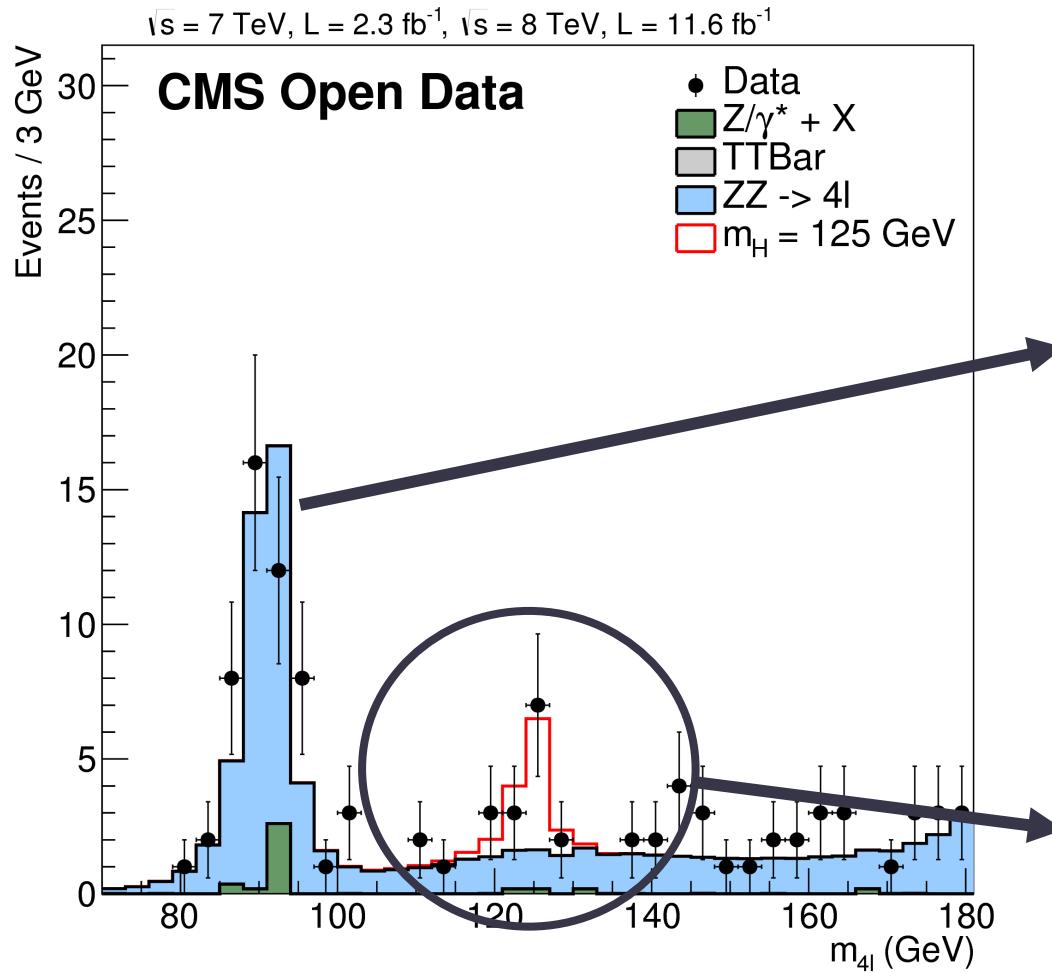
# Higgs discovery: simplified reimplementation



# Background

This data set is too small to say anything for certain, but it isn't too far off from actual analysis results.

# Higgs discovery: simplified reimplementation

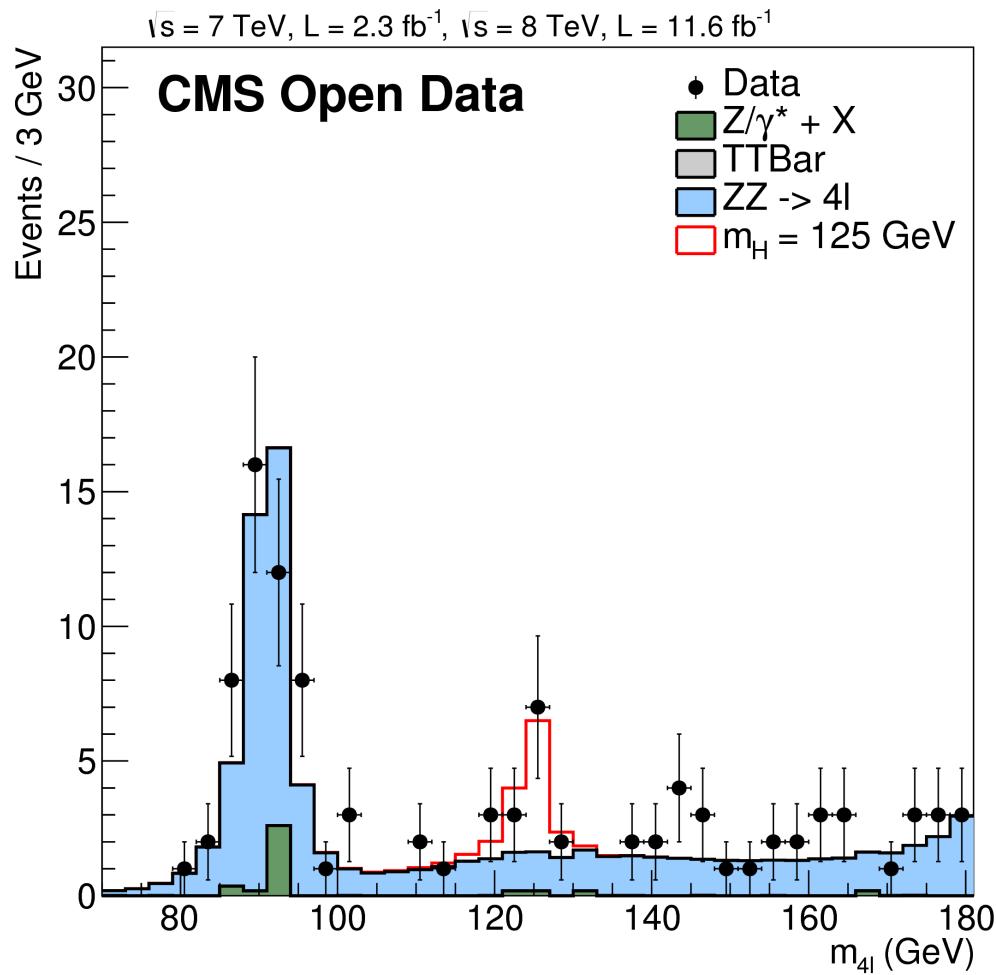


Background

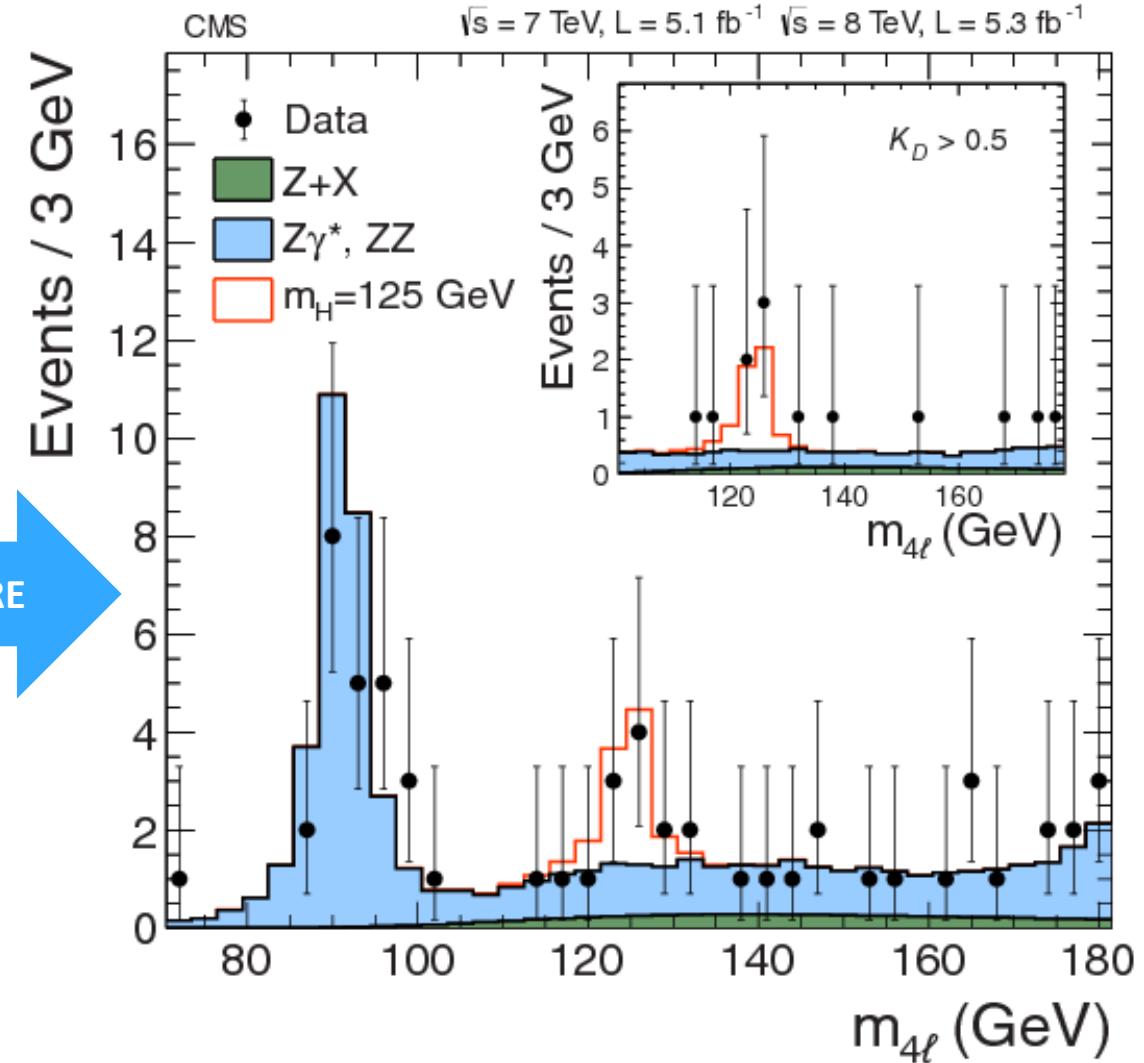
Signal

This data set is too small to say anything for certain, but it isn't too far off from actual analysis results.

# Higgs discovery: simplified reimplementation



COMPARE



## Summary



CERN  
ANALYSIS PRESERVATION

H $\rightarrow$ 4l decay analysis example is fully reproducible.

Add documentation of previous examples to CAP

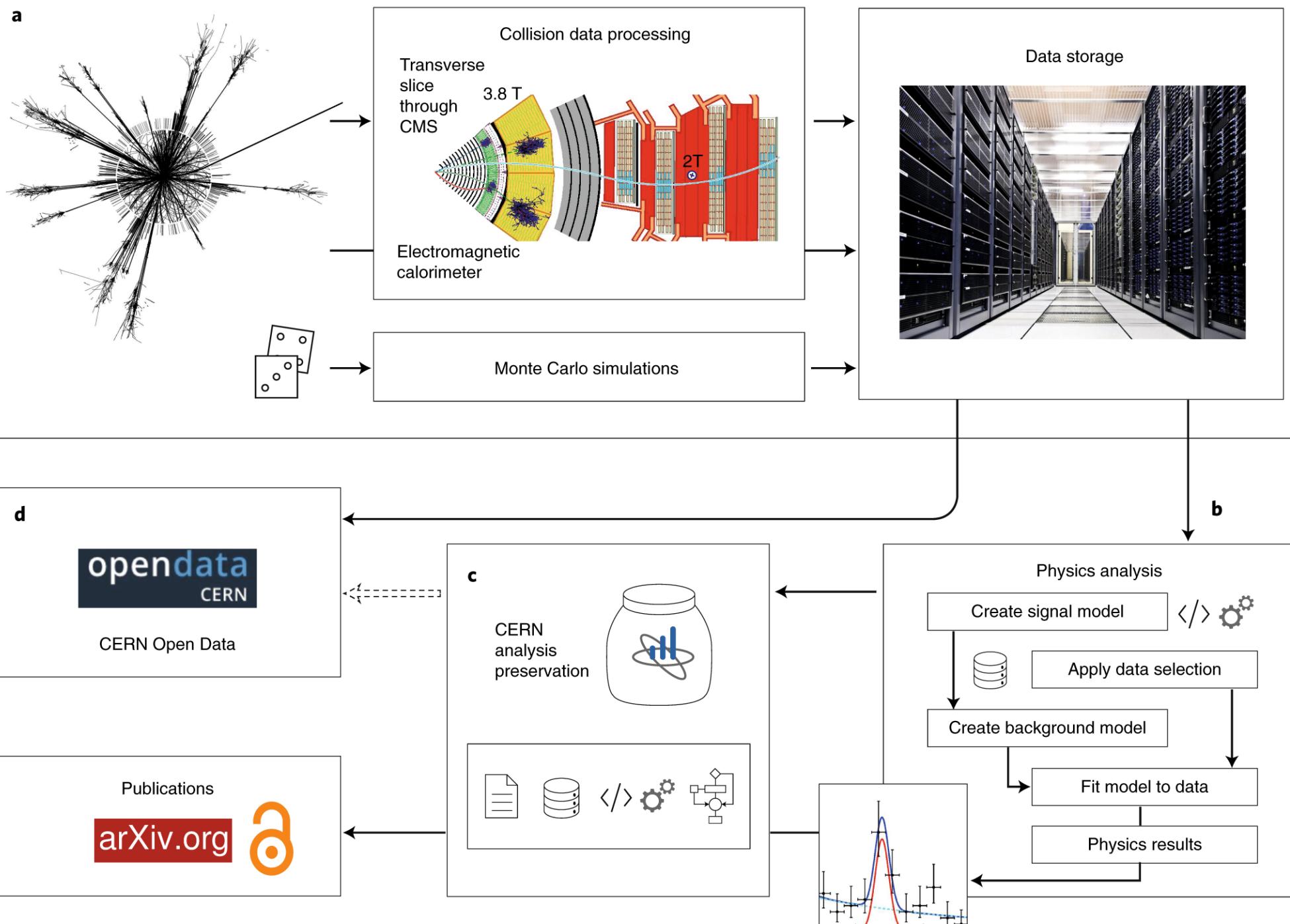
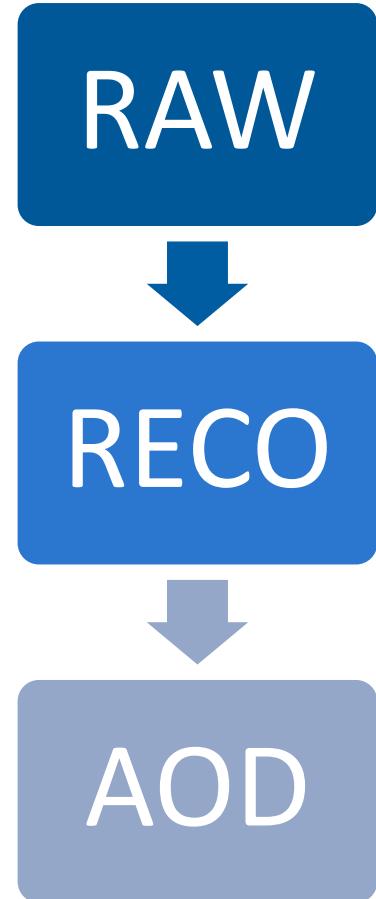
## Next Steps

- ❖ UPRM CMS Research Group: Analysis for Supersymmetry
- ❖ Top quark mass measurement from b-jet energy spectrum analysis example

# Backup



Reprocessing AOD from 2010 -  
2012 RAW samples



# Selected RAW Samples\*



## 2010 DATASETS

/MinimumBias/Run2010B-v1/RAW  
/Electron /Run2010B-v1/RAW  
/Mu /Run2010B-v1/RAW  
/Jet /Run2010B-v1/RAW



## 2011 DATASETS

/DoubleElectron/Run2011A-v1/RAW  
/SingleElectron /Run2011A-v1/RAW  
/DoubleMu /Run2011A-v1/RAW  
/SingleMu /Run2011A-v1/RAW  
/Jet /Run2011A-v1/RAW

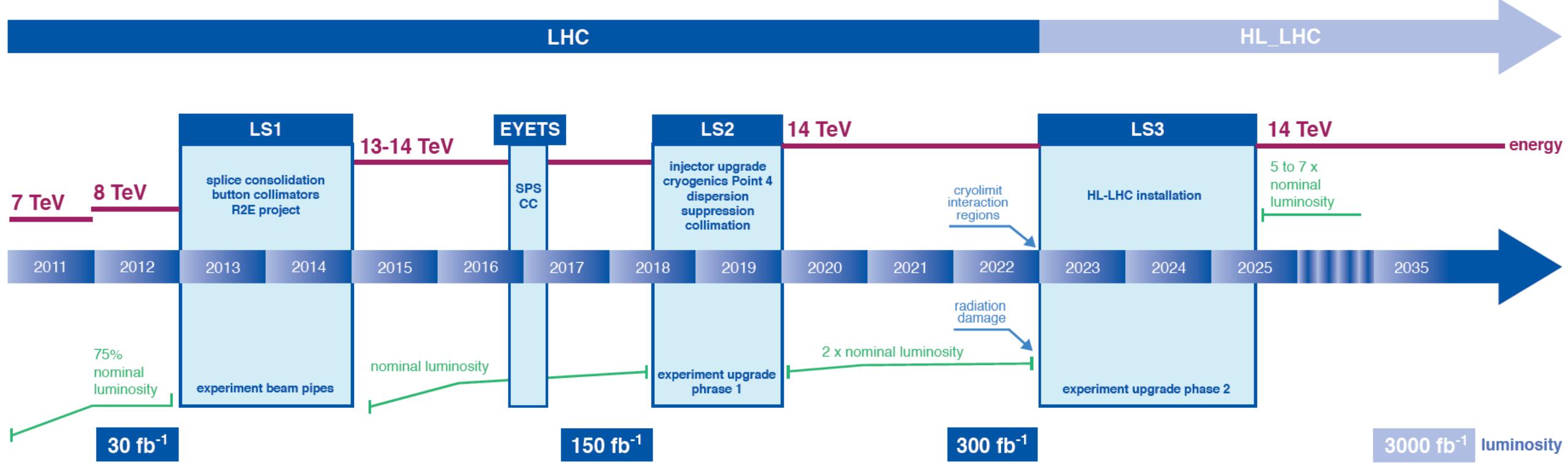


## 2012 DATASETS

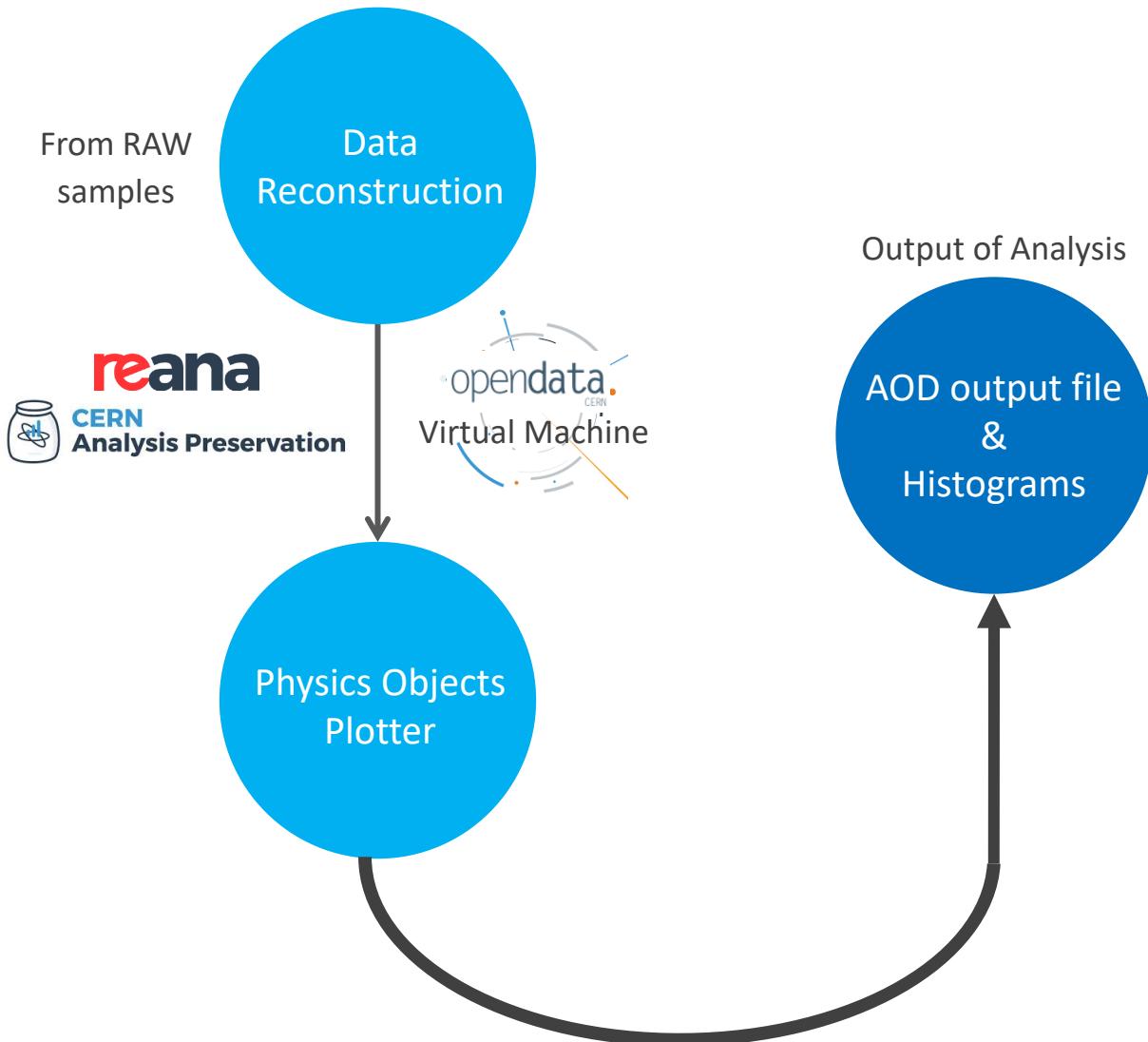
/DoubleElectron/Run2012B-v1/RAW  
/SingleElectron /Run2012B-v1/RAW  
/DoubleMuParked/Run2012B-v1/RAW  
/SingleMu /Run2012B-v1/RAW  
/JetHT/Run2012B-v1/RAW

\*TO BE RELEASED TO OPEN DATA PORTAL THIS YEAR

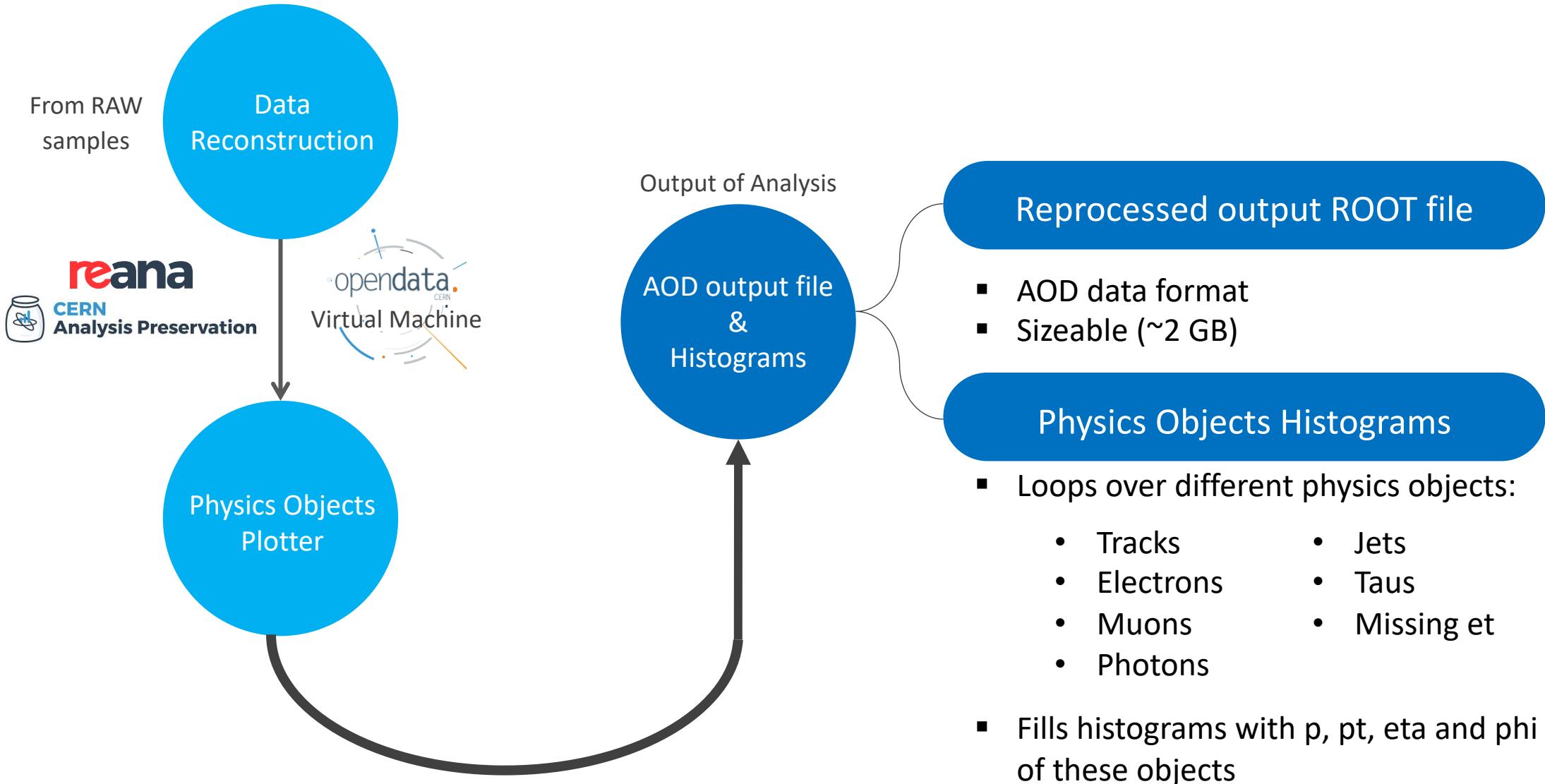
# LHC / HL-LHC Plan



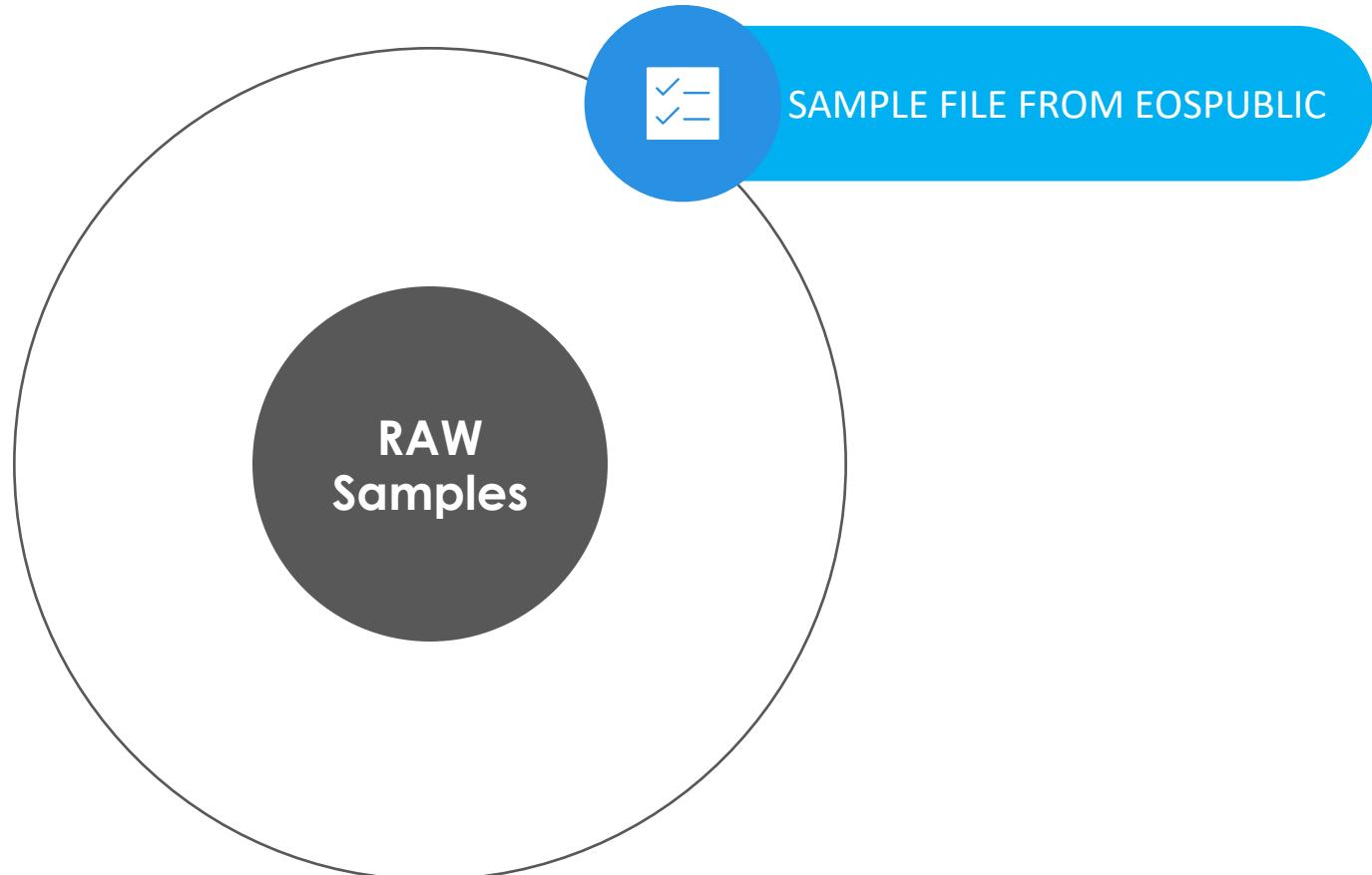
# Project Analysis



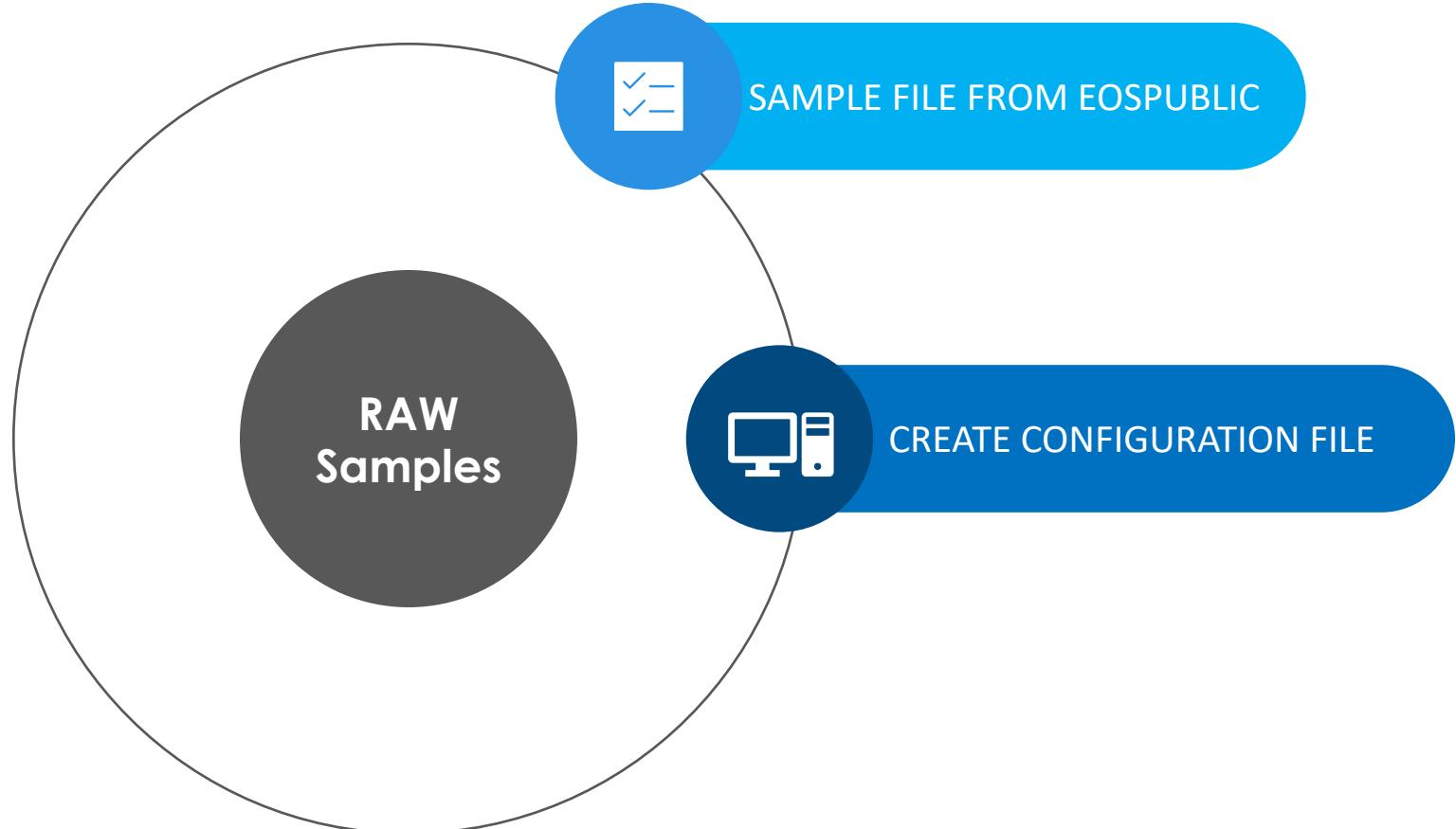
# Project Analysis



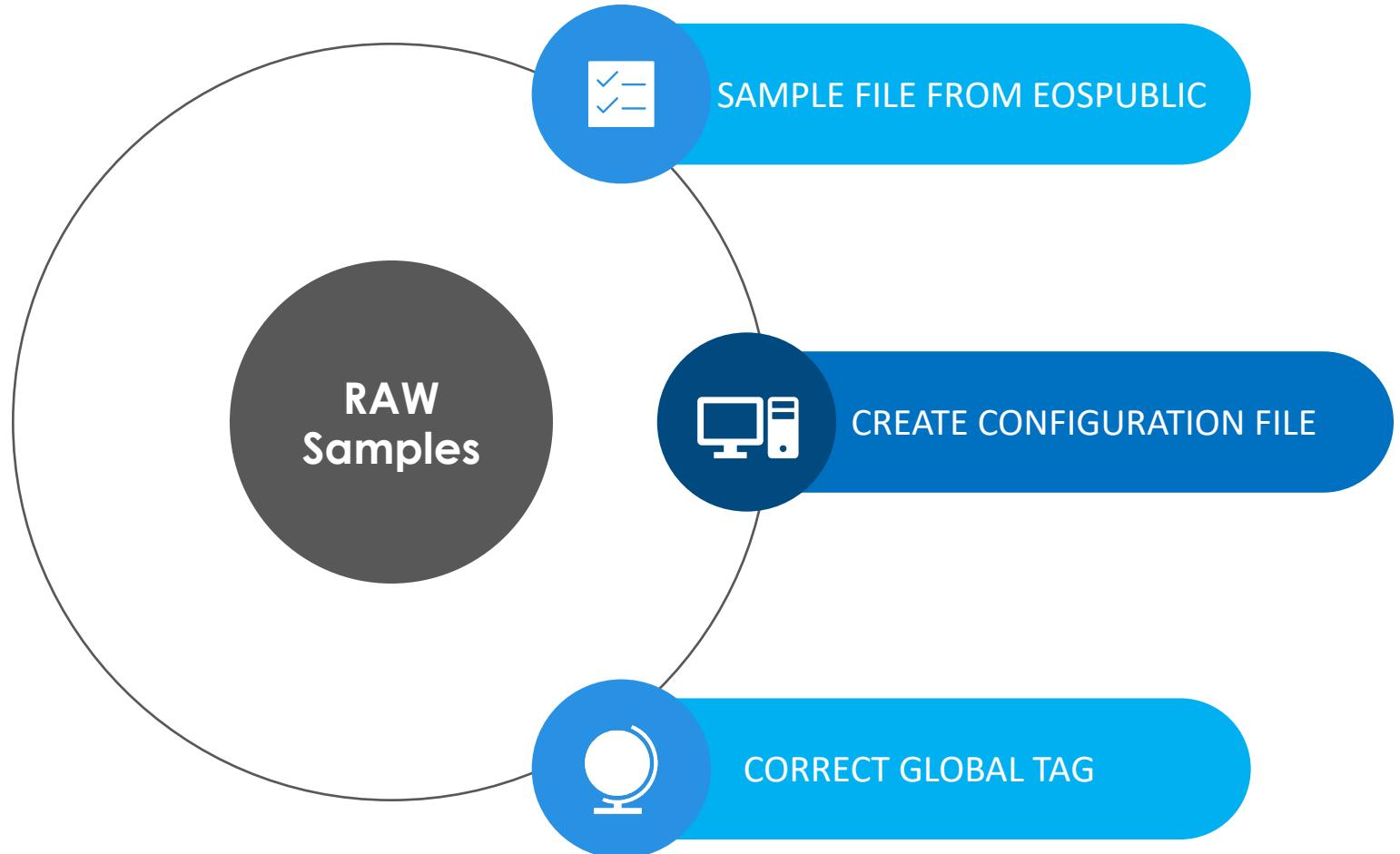
# Data Reconstruction Process



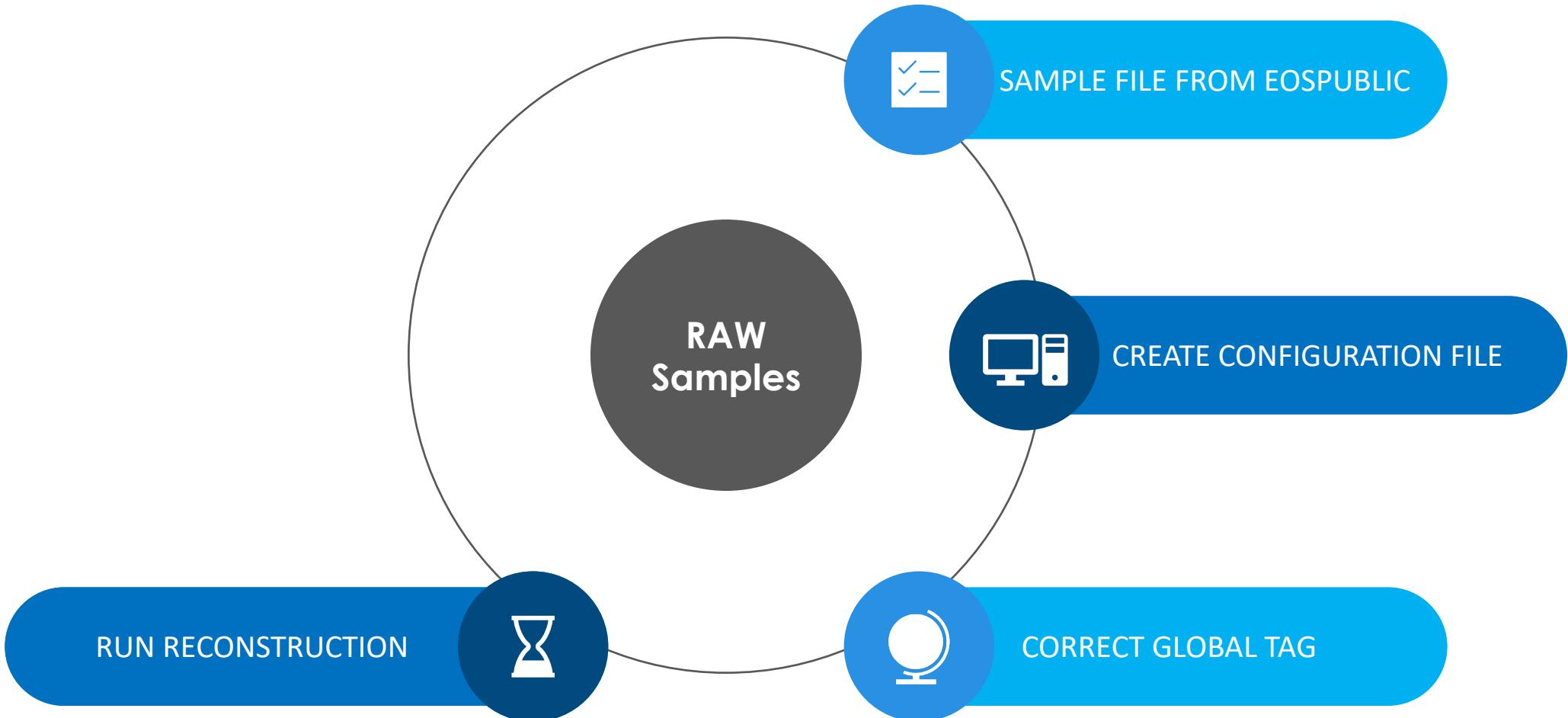
# Data Reconstruction Process



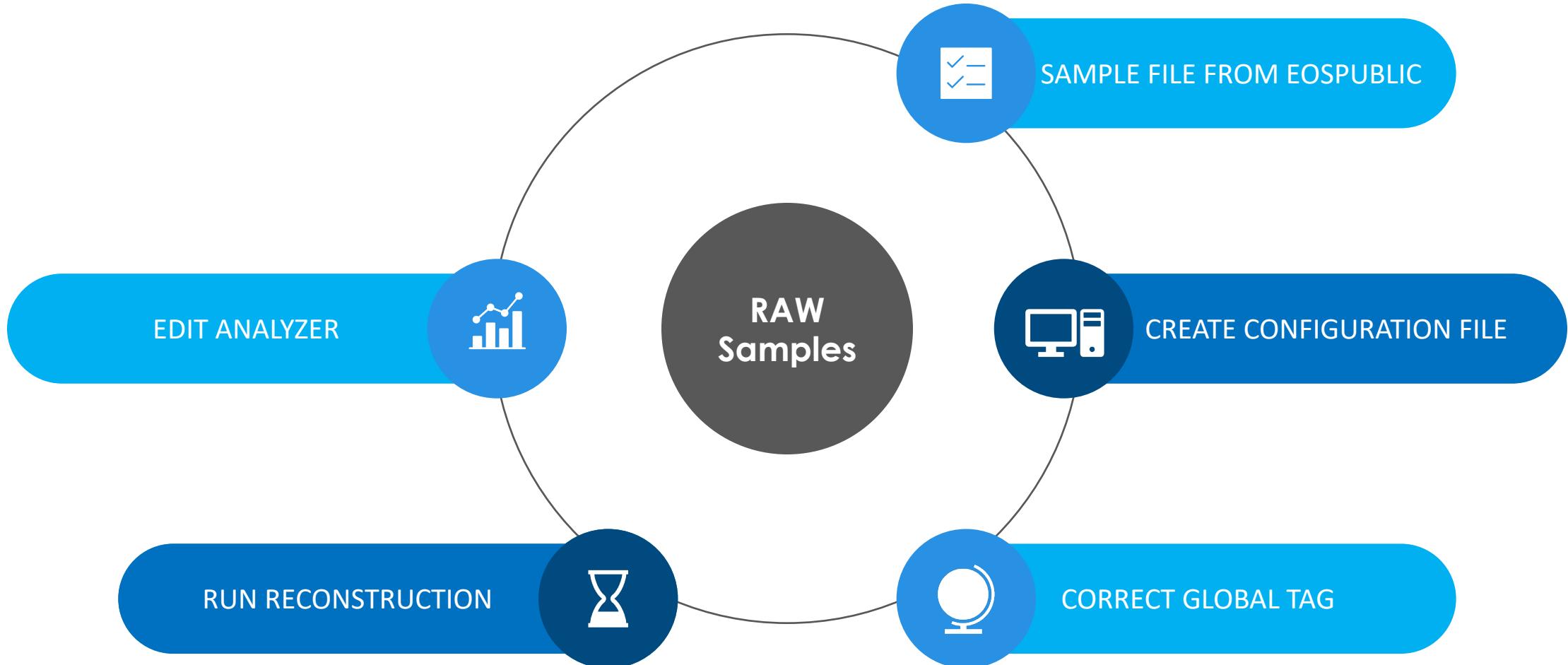
# Data Reconstruction Process



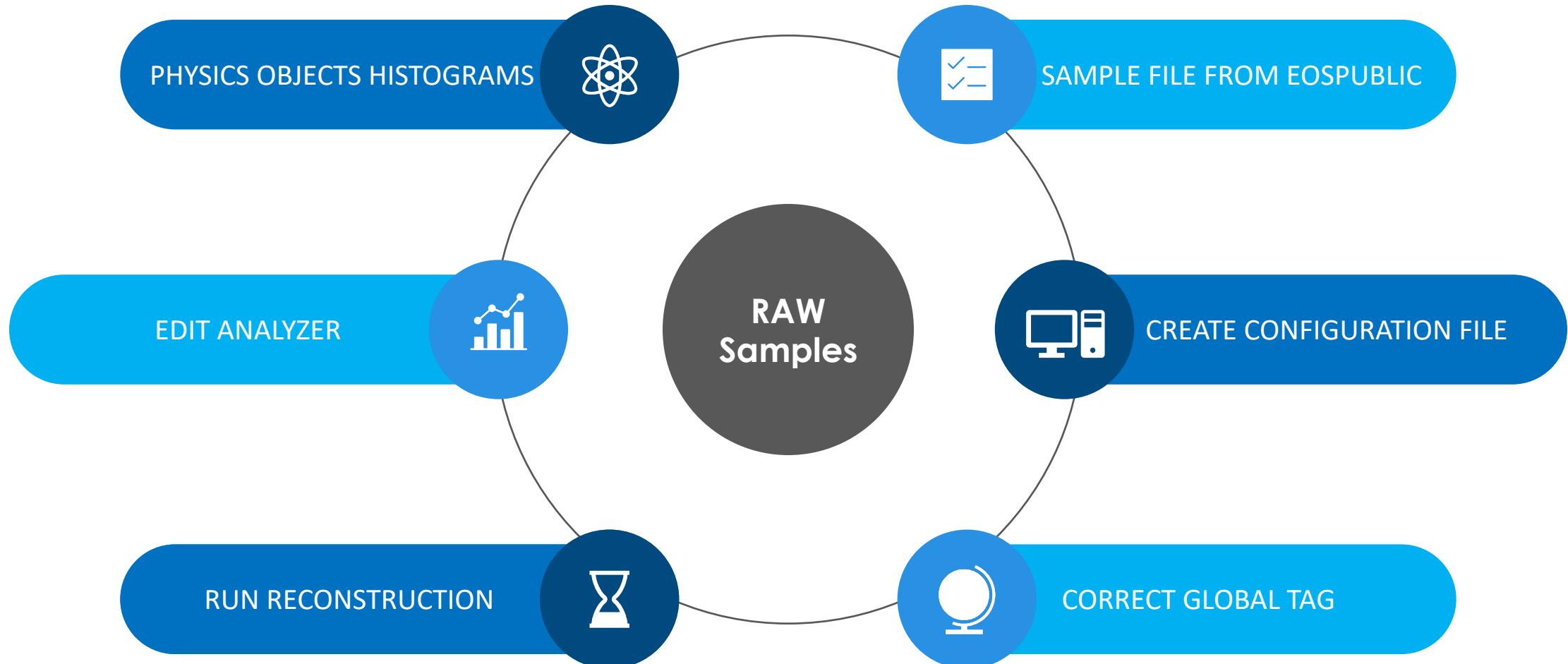
# Data Reconstruction Process



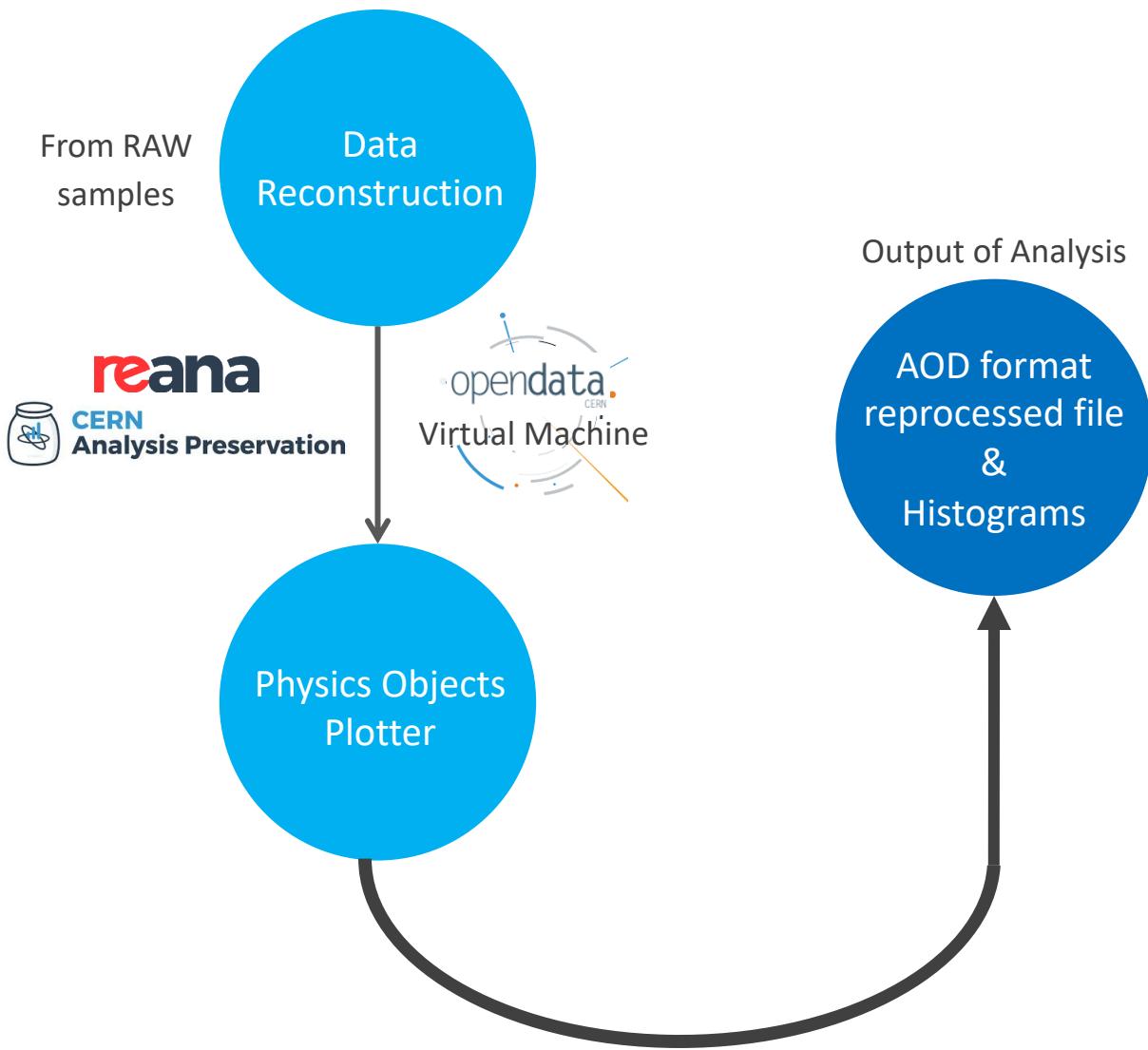
# Data Reconstruction Process



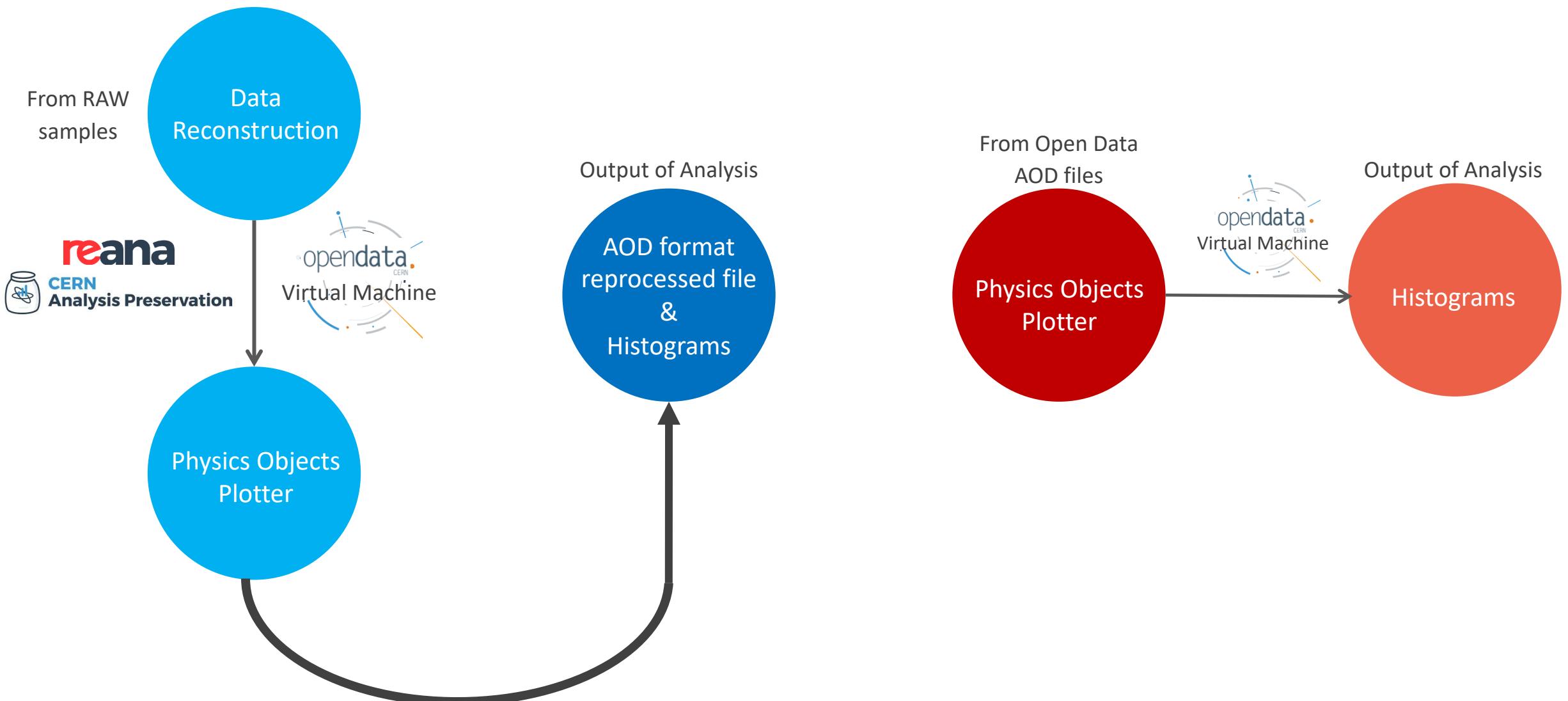
# Data Reconstruction Process



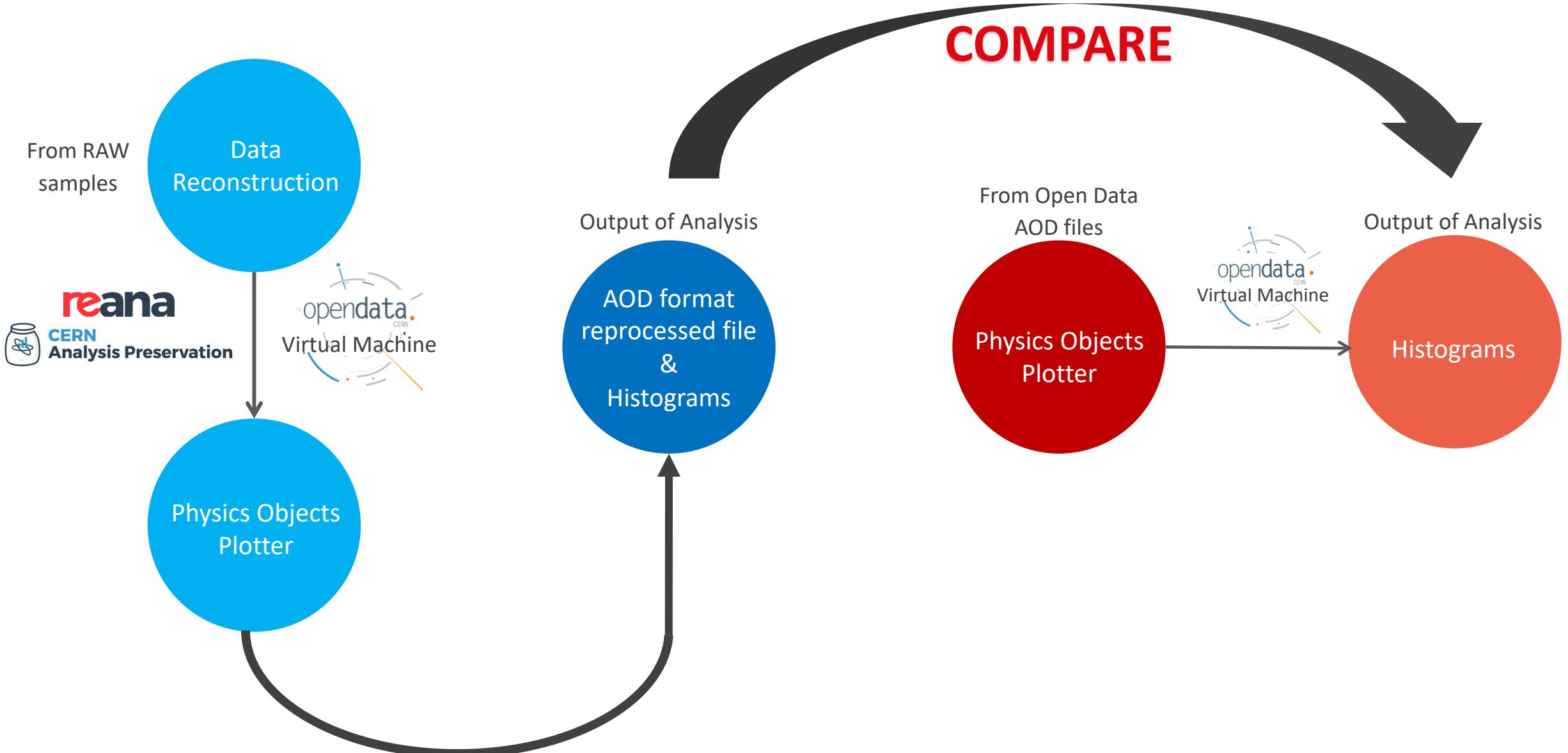
# Project Analysis



# Project Analysis

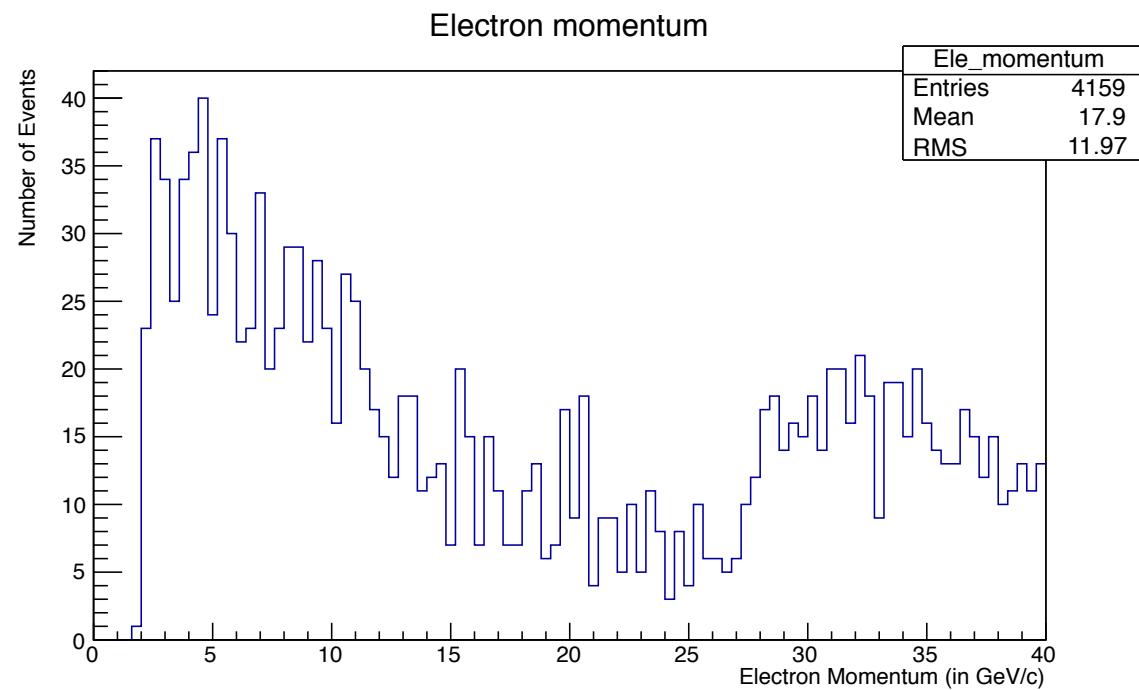


# Project Analysis



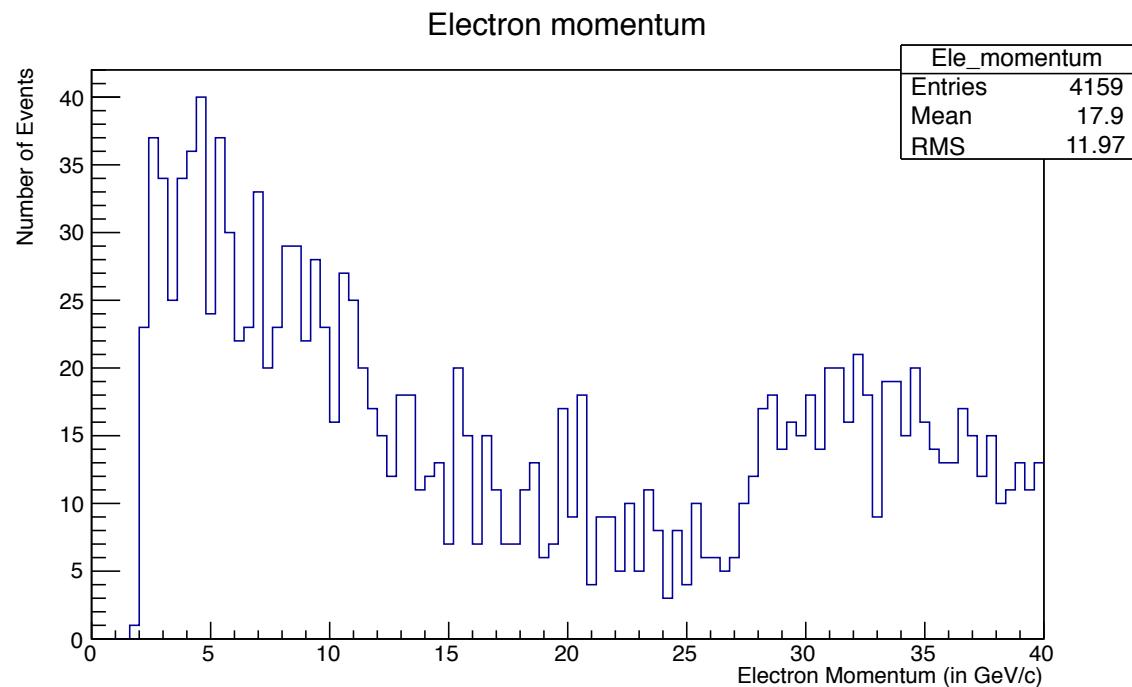
# Run2011A: SingleElectron

Reconstruction from  
RAW samples

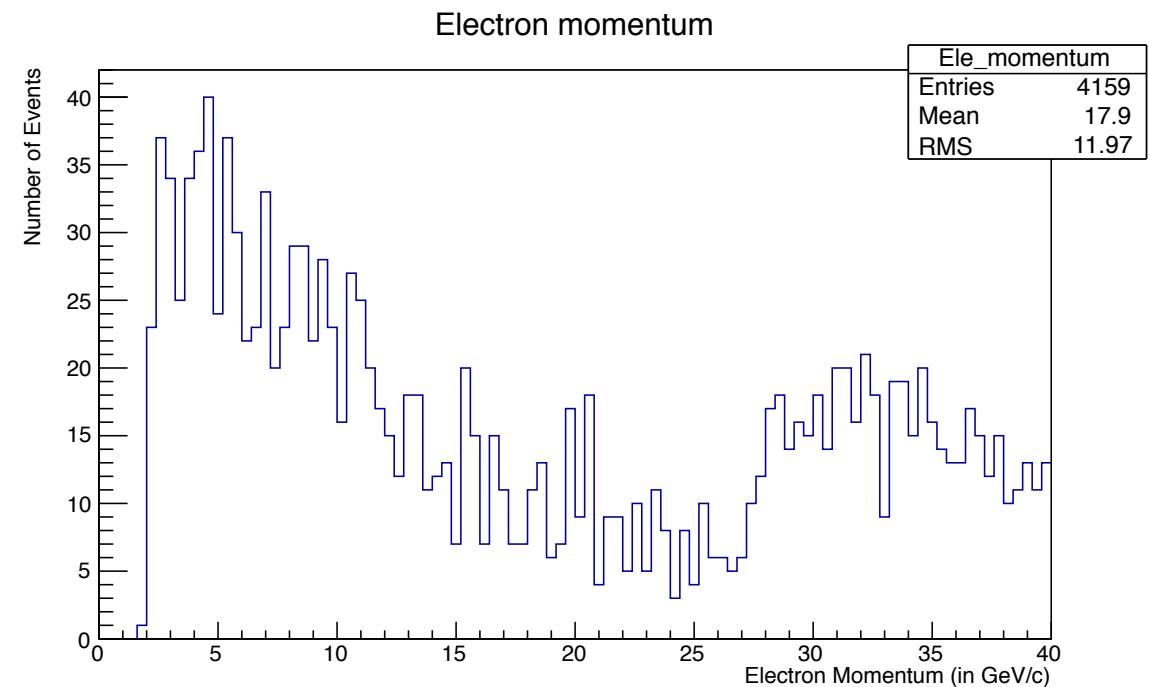


# Run2011A: SingleElectron

Reconstruction from  
RAW samples

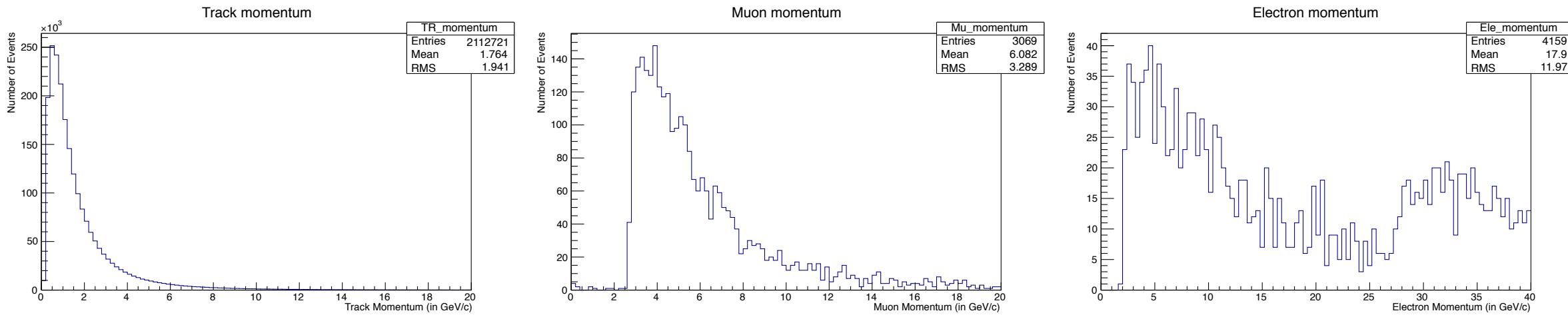


CMS Open Data  
AOD files

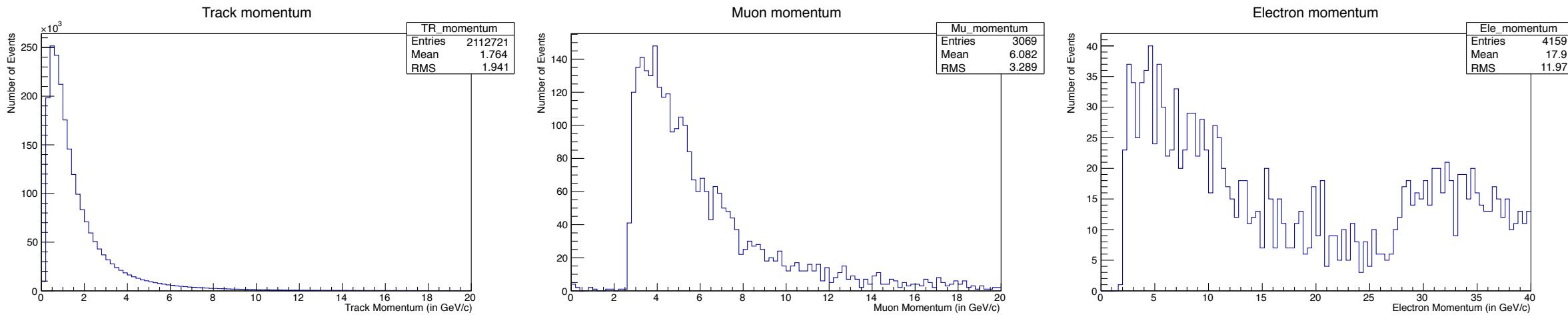


## Reconstruction from RAW samples

# Run2011A: SingleElectron

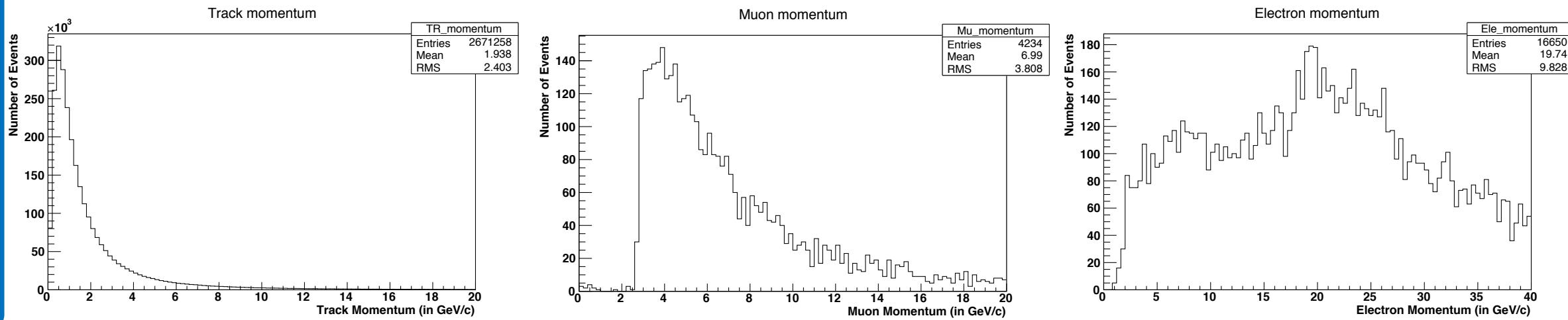


## CMS Open Data AOD files

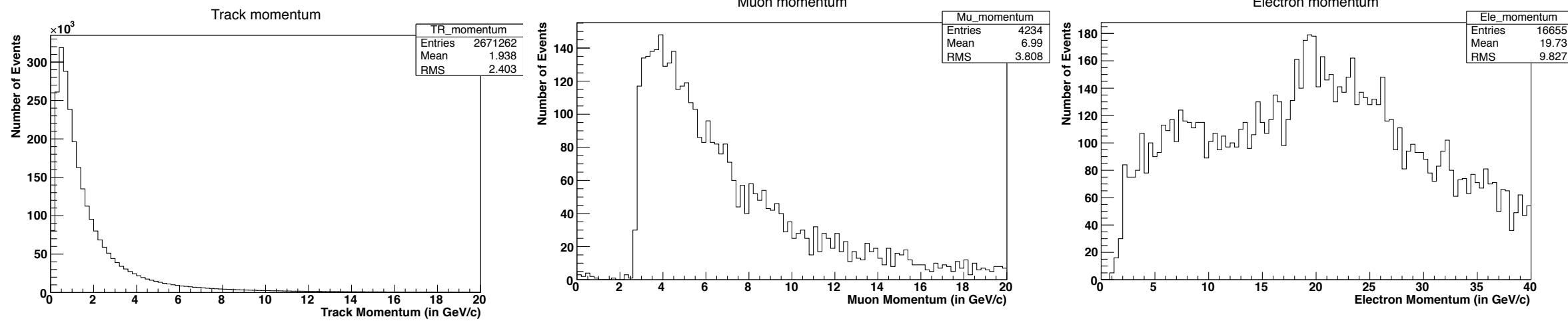


## Reconstruction from RAW samples

# Run2010B: Electron



## CMS Open Data AOD files



## Summary



CERN  
ANALYSIS PRESERVATION

H $\rightarrow$ 4l decay analysis example is fully reproducible.

Add documentation of previous examples to CAP

All RAW samples were reconstructed successfully,  
and have one-to-one match with the original AOD.

## Next Steps

- ❖ UPRM CMS Research Group: Analysis for Supersymmetry
- ❖ Top quark mass measurement from b-jet energy spectrum analysis example