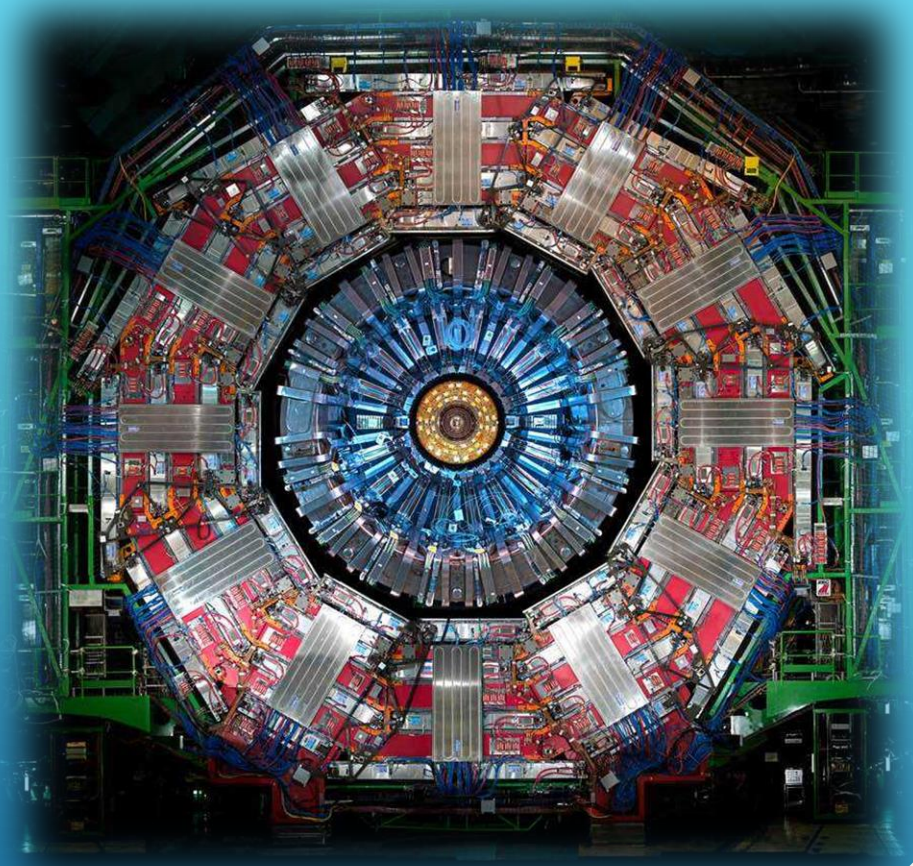
EXPERIMENT PERSPECTIVE: CMS



Lara Lloret Iglesias

IFCA

On behalf of the DPOA group

ANALYSIS PRESERVATION

Three different approaches (common to all experiments though priorities may and will differ)

Analysis Description → **Make information searchable**

- Being able to search the analysis by CADI number, by final state, by pt cut, by trigger, by samples used…

Analysis Capture → **Preserve assets someplace safe**

- Nothing fancy: just storing the main code and the scripts somewhere safe with a README with **simple instructions** explaining how to run each of them, what they take as input/output , in which order run the code…

- A place to store the final ntuples

Analysis Reuse (or Reproducibility) → **Rerun it with same/modified inputs**

The most tricky part…

The full computational workflow steps

Not easy in general, but even less when people have loosely connected set of scripts (almost always ☺)

Intermediate steps possible → preserve final ntuples (CAP), make final plots (REANA)…

ANALYSIS PRESERVATION

Right now:

Trying to make the CAP **versatile enough** to cover, to some extent, the three levels

- Searchable analysis

- Basic preservation for code/ntuples

- Some "basic" placeholders trying to include as much as possible all the steps/inputs used during the analysis workflow: control regions, efficiencies, scale factors…

Also from the visual point of view:

- Default version containing the **two first levels** → more user friendly

- Possibility to go to an extended version including placeholders for many more details → **first test** for a total analysis reuse

Probably we will have to go from a light approach (final plots, final ntuples) to a more sophisticated one (whole workflow)

Not all analysis will be interested in the reproducibility part → Maybe for really important analysis with a lot of visibility outside? (i.e Higgs)

Summer students:

For testing the eventual CAP/ReANA integration at least for the open data examples