

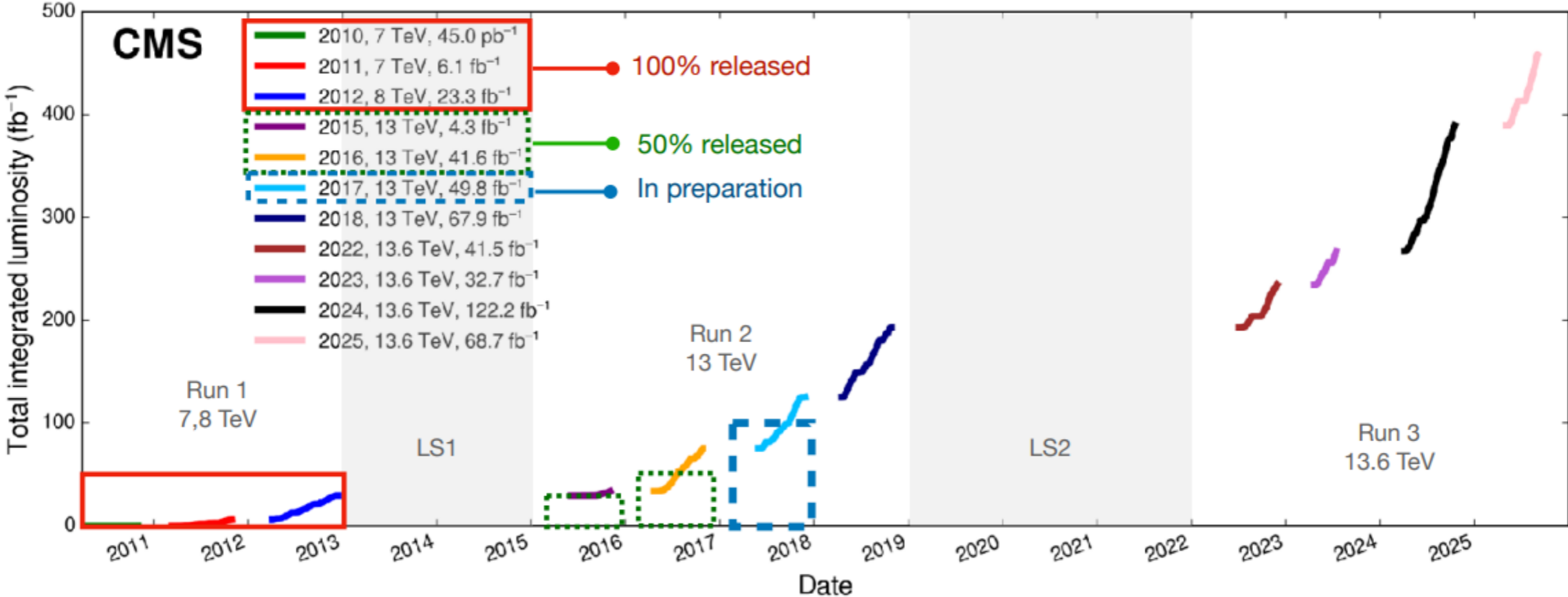


Status of CMS open data & tools

Oz & Abhijith

Current state of CMS open data

- A decent chunk of Run-2 CMS data is available on open data portal
 - More incoming



From Matthew. B

Data formats available

- **AOD**: largest data format [Run I + MC only]
 - Full detector output with very low-level information
- **MiniAOD**: smaller data format derived from AOD
 - Physics objects + PF candidates + tracks + vertices etc ...
 - Contains reduced collection of detector hits
- **NanoAOD**: flat, ROOT-based ntuples
 - Physics objects
 - Analysts can use standard python frameworks and tools such as Coffea, RDataFrame, uproot, awkward, ...

Data formats available

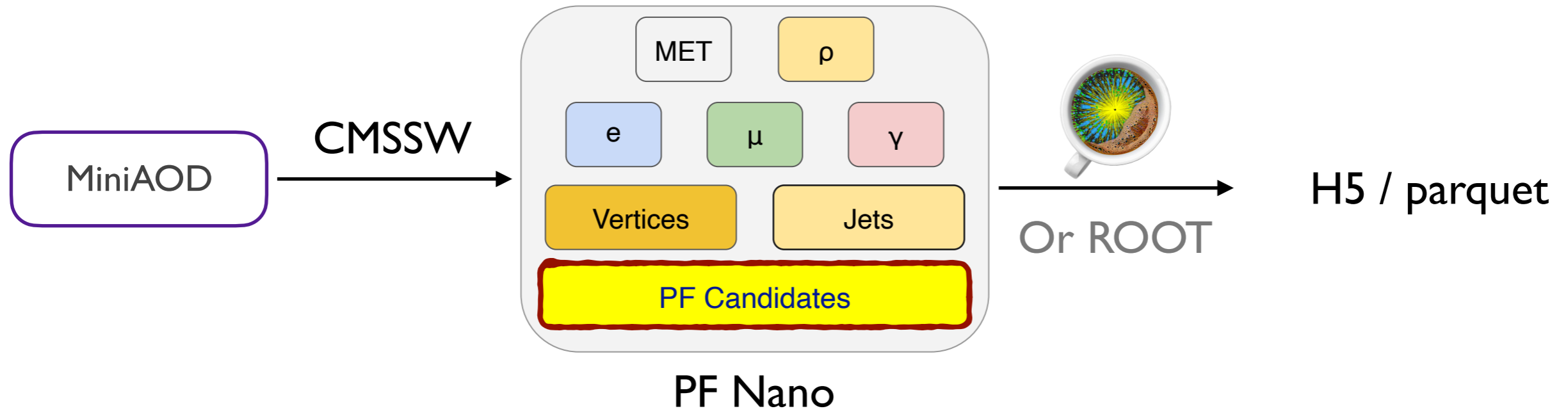
- **AOD**: largest data format [Run I + MC only]
 - Full detector output with very low-level information
- **MiniAOD**: smaller data format derived from AOD
 - Physics objects + PF candidates + tracks + vertices etc ...
 - Contains reduced collection of detector hits
- **NanoAOD**: flat, ROOT-based ntuples
 - Physics objects
 - Analysts can use standard python frameworks and tools such as Coffea, RDataFrame, uproot + DASK
- **PF NanoAOD**: flat, ROOT-based ntuples
 - Physics objects + PF candidates
 - Can use Coffea schema to convert to Awakward arrays

Needs to be run
by the users

Code in opendata
portal

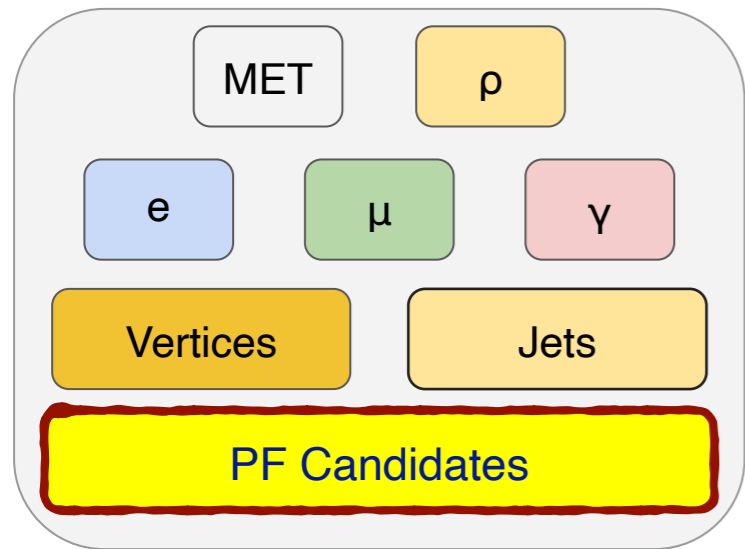
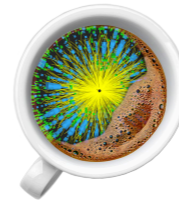
Workflow

- Publicly available code for make these data conversion

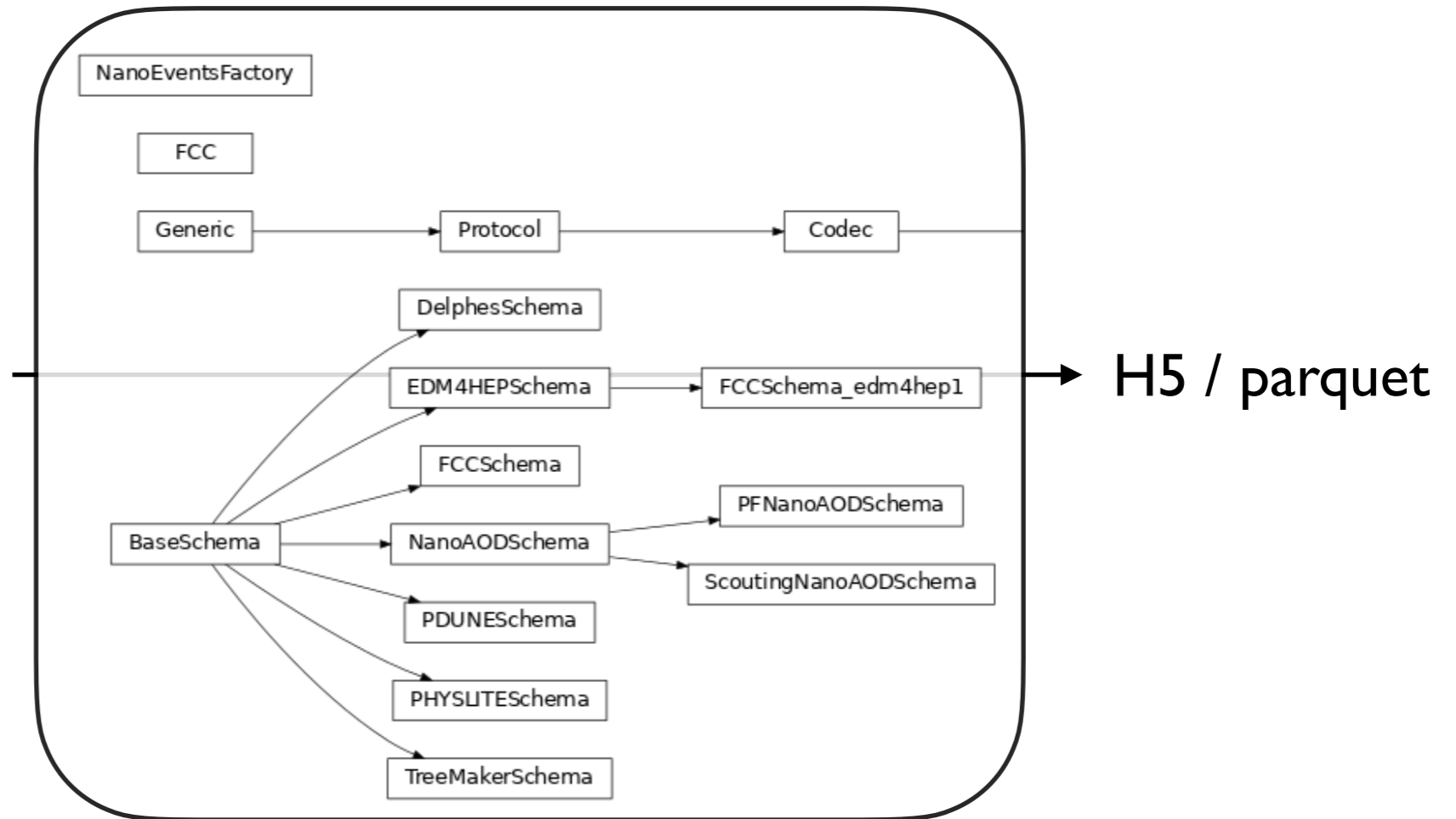


Workflow

- Publicly available code for make these data conversion



PF Nano



Double-jagged PF candidate collection

Workflow - Extending to low level

- Publicly available code for make these data conversion

