

# Data flow and Event Data Model

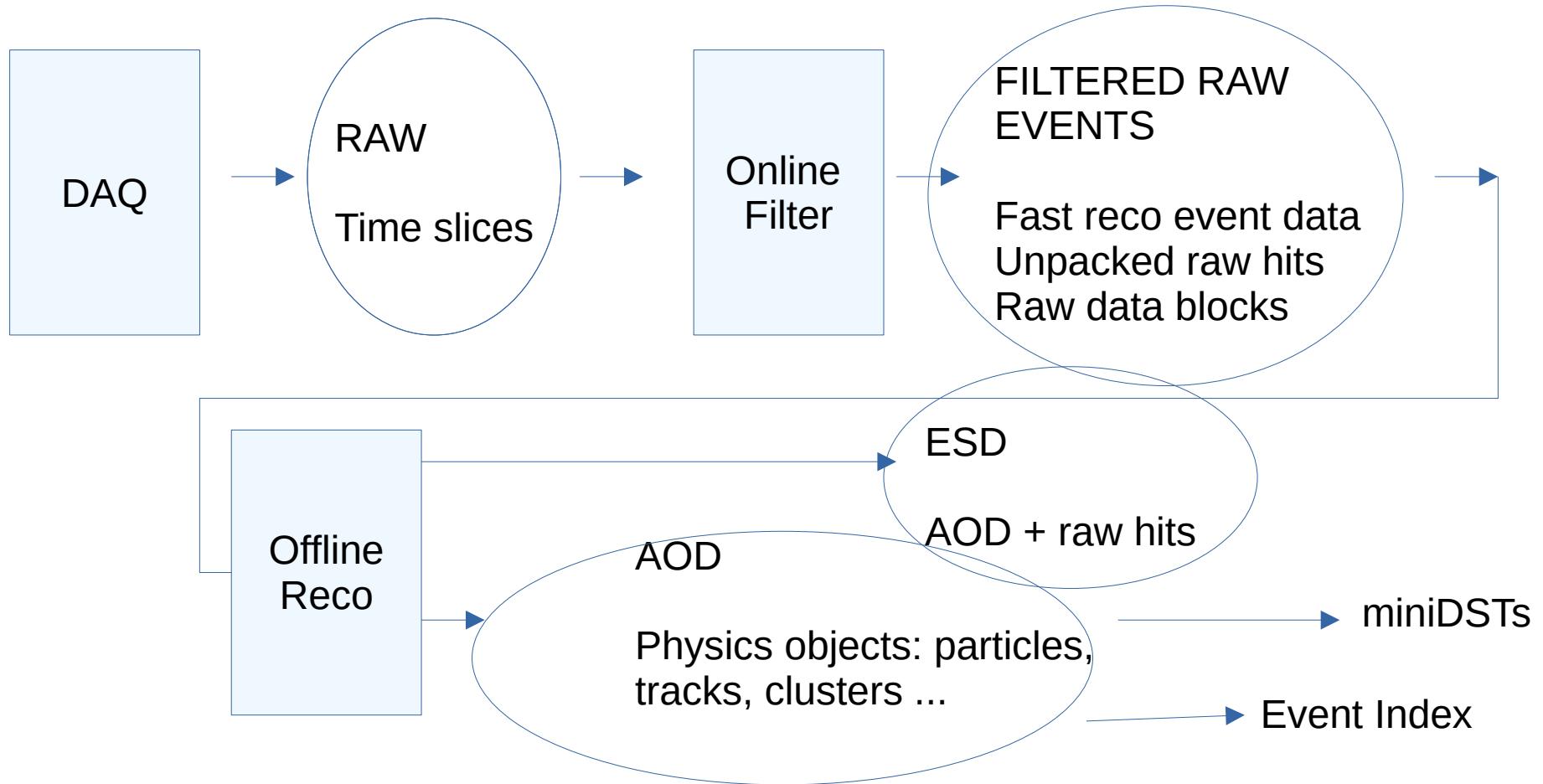
Alexey Zhemchugov, Igor Denisenko

05.10.2022

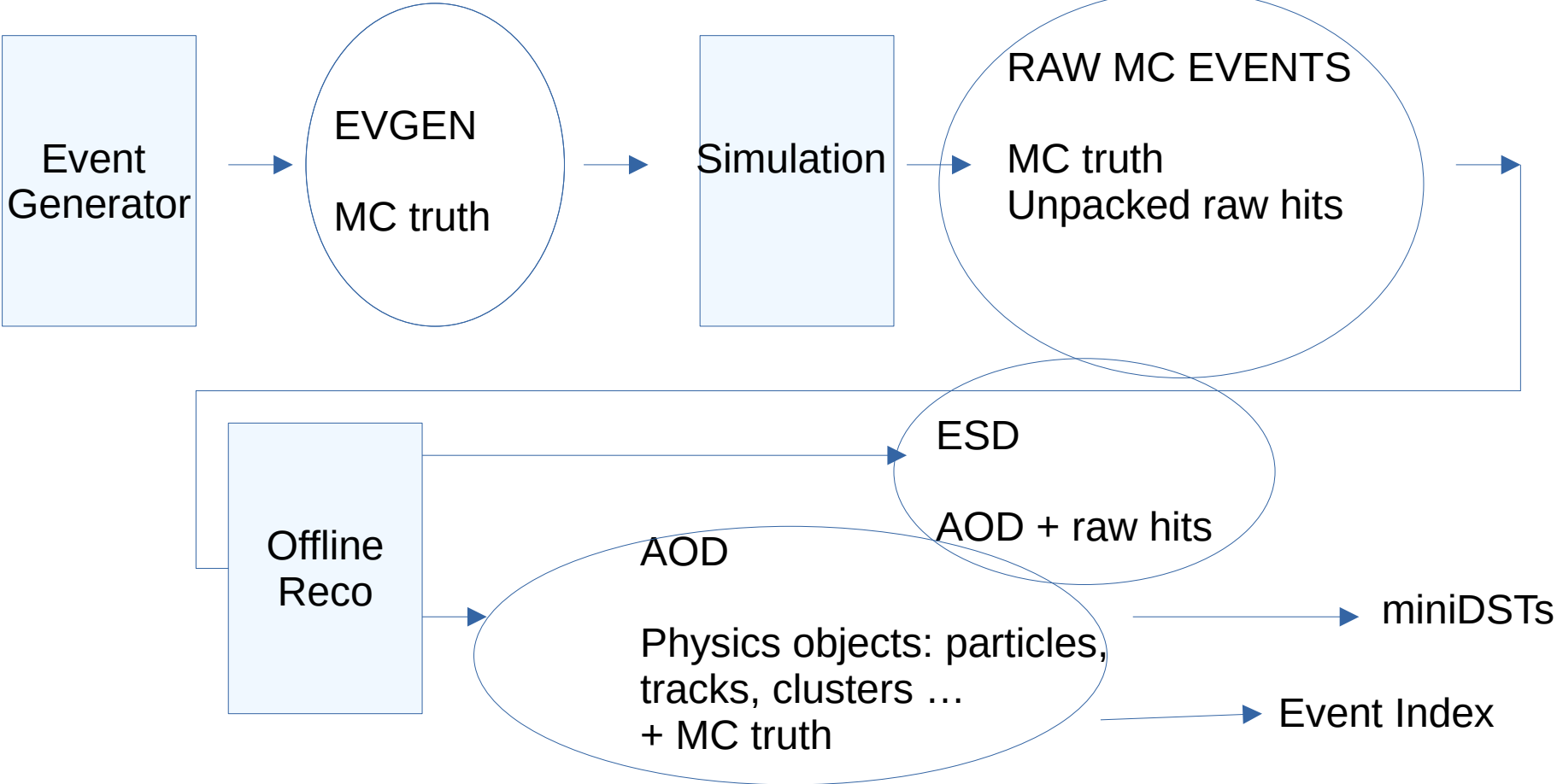
# Data flow

- Real data
- MC data
- ML training

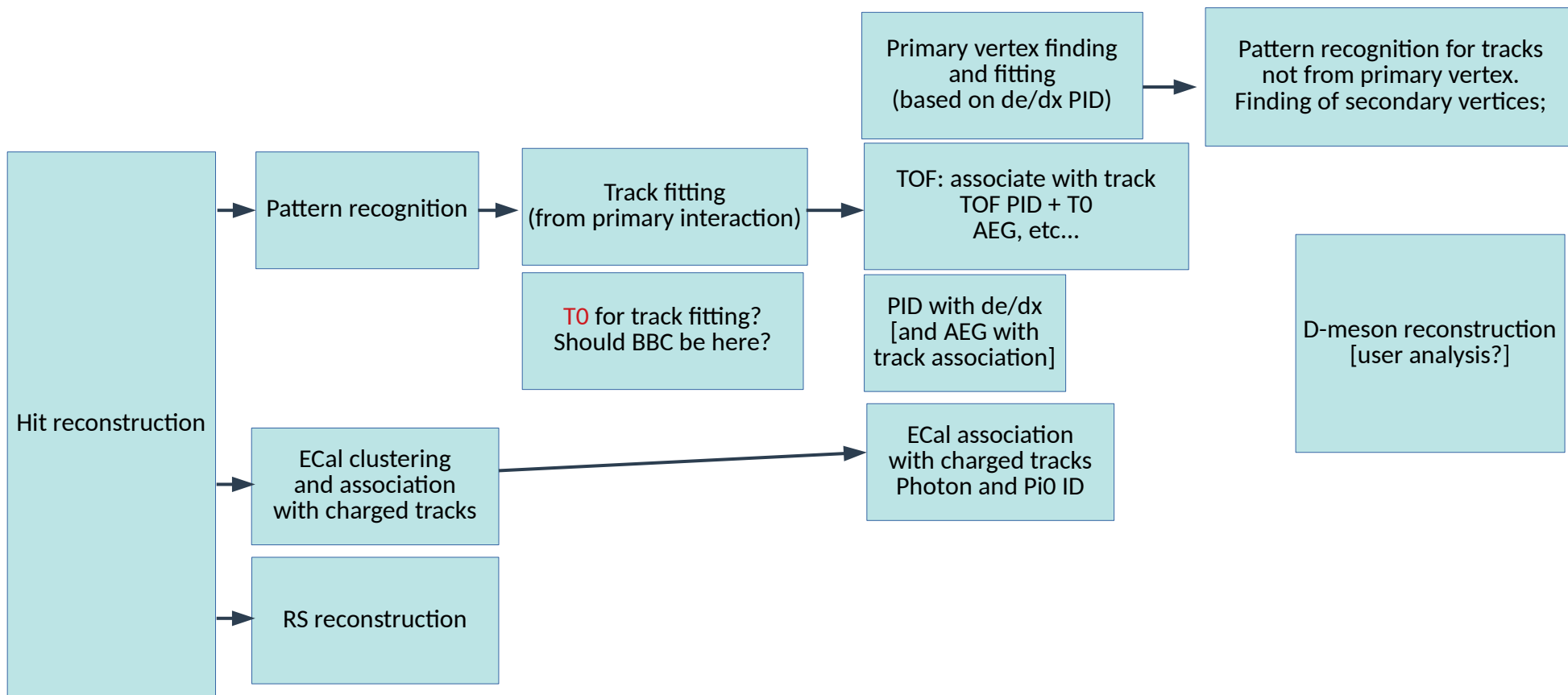
# Real data



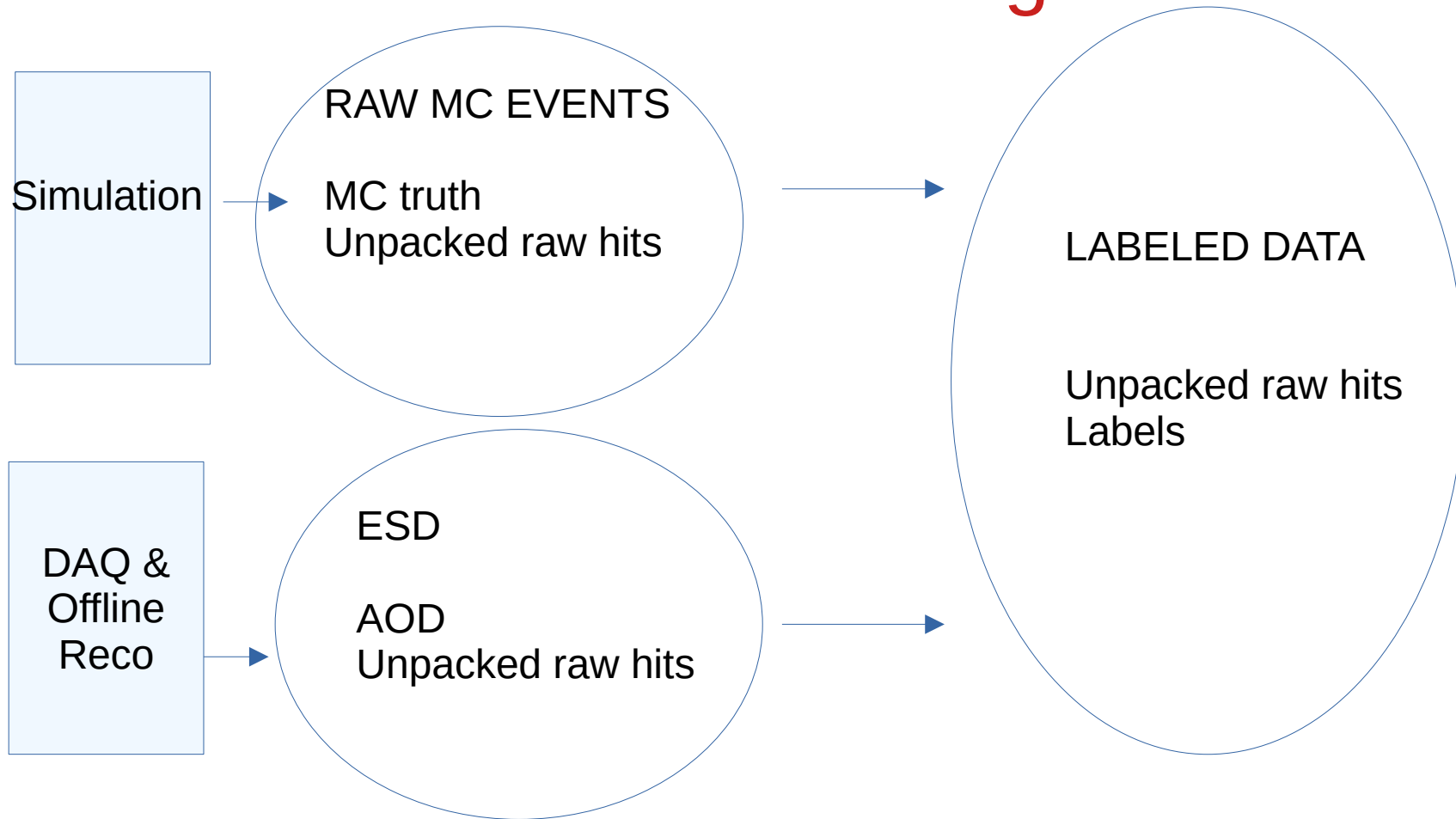
# MC data



# Reconstruction chain: more details



# ML training

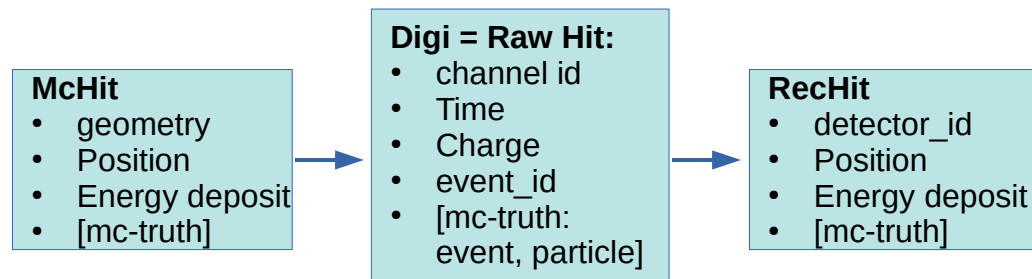


# Data types

- *RAW TIME SLICES (determined by the DAQ)*
- FILTERED RAW EVENTS
- AOD
- ESD (= AOD + hits)
- EVGEN (=HepMC?)
- RAW MC EVENTS
- LABELED DATA

*We need the definition of each data type and transformations to process one data format to another*

# Draft AOD/ESD model



## MCParticle

- id
- status
- PDG\_code
- prod\_vtx
- prod\_time
- initial\_momentum
- mother[]
- daughter\_list[]

## MCVertex

- id
- position
- MCParticles[]

## Track:

- charge
- *hits*[]
- momentum\_in\_vertex [e, mu, pi, K, p, d?] and covariance
- chi2[]
- ndf[]
- vertex\_id
- PID info [de/dx, TOF, AEC]
- association to BBC, ECal, RS
- track states[] in the last and the first track point for extrapolation
- [MC particle id]

## Primary/Secondary Vertex:

- id
- is\_primary
- position + covariance
- tracks[]
- chi2
- ndf

## ECal cluster:

- energy
- position
- barrel/endcap flag
- *hits*[]
- [MC particle id]

## RS segment:

- extrapolated length
- number of hits in a cone
- energy\_estimate
- ...
- [MC particle id]

## PID info:

### De/dx

- prob(particle\_type)?
- *hits\_ids*[]

### TOF

- prob(particle\_type)
- cell\_id
- track extrapolation pos in a cell

### AEC

- has\_fired
- cell\_id
- track extrapolation pos in cell

### ECal?

- prob

### RS

- prob

**Every extra byte will become one extra Terabyte per year!**