INSPECTING EDM4HEP FILES

Juraj Smieško

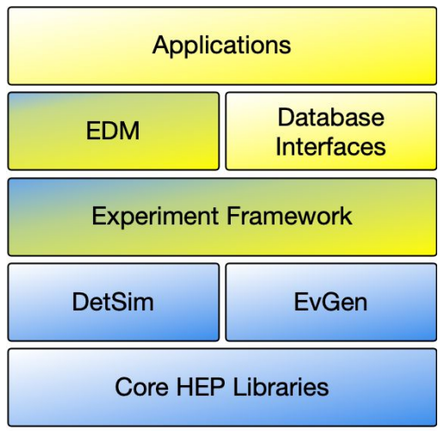
CERN

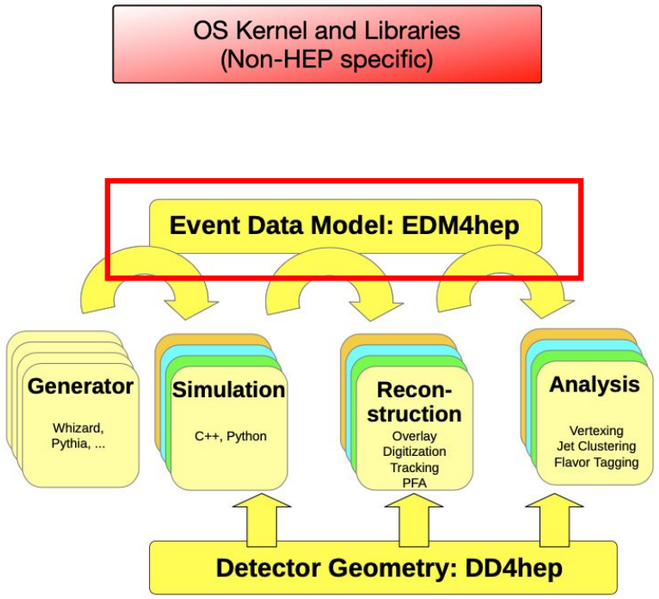
FCC Software Meeting

CERN, 25 Sep 2023

KEY4HEP

- Set of common software packages, tools, and standards for different Detector concepts
- Common for FCC, CLIC/ILC, CEPC, EIC, …
- Individual participants can mix and match their stack
- Main ingredients:
  - Data processing framework: Gaudi
  - Event data model: EDM4hep
  - Detector description: DD4hep
  - Software distribution: Spack

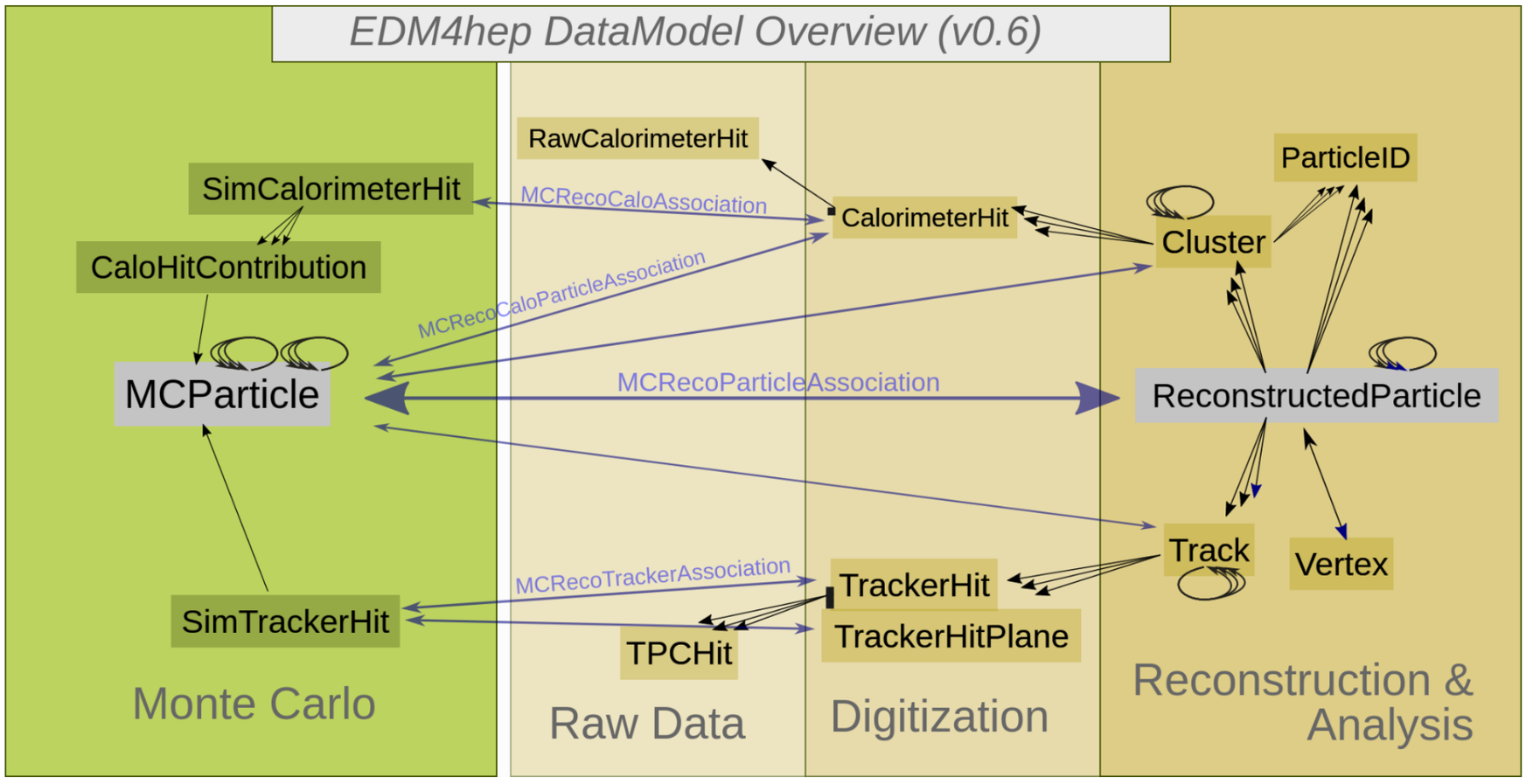




EDM4HEP I.

Describes event data in a set of standard objects and relationships between them

- Specification in a single YAML file
- Strives to be minimal
- Based on LCIO and FCC-edm



EDM4HEP II.

Example object:

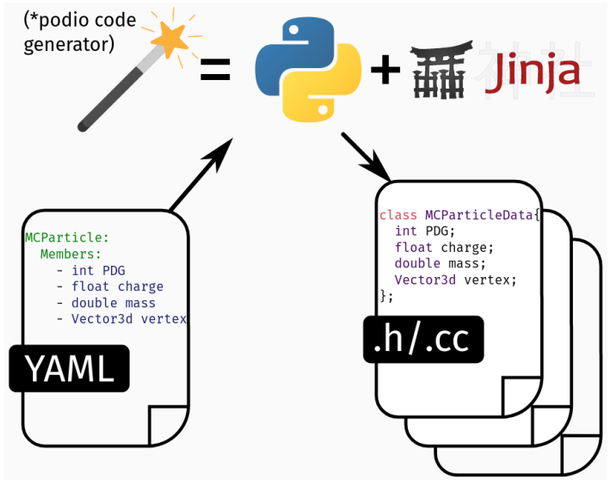
```
#------------- SimCalorimeterHit
edm4hep::SimCalorimeterHit:
  Description: "Simulated calorimeter hit"
  Author: "F.Gaede, DESY"
  Members:
  - uint64_t cellID       //ID of the sensor that created this hit
  - float energy                    //energy of the hit in [GeV].
  - edm4hep::Vector3f position      //position of the hit in world coordinates in [mm].
OneToManyRelations:
  - edm4hep::CaloHitContribution contributions  //Monte Carlo step contribution - parallel
```

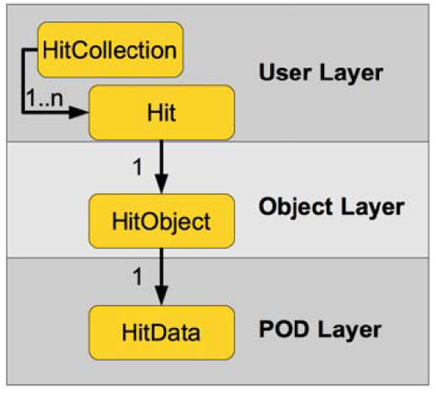
- Current version: `v0.10.0`
- Warning: Policies how to save event data into datamodel in flux
- Objects can be extended / new created
- Bi-weekly discussion: Indico

PODIO

Generates Event Data Model and serves as I/O Layer

- Generates EDM from YAML files
- I/O machinery consists of three layers
  - POD Layer - arrays of the actual data structures
  - Object Layer - handles the relations
  - User Layer - handles to the EDM objects
- Supports multiple backends:
  - ROOT, SIO
- Current version: `0.17.0`

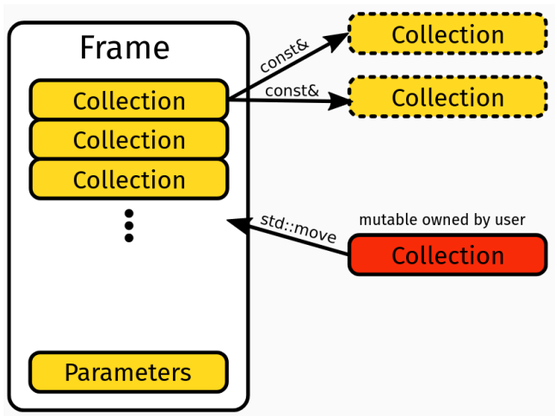




PODIO: RECENT CHANGES

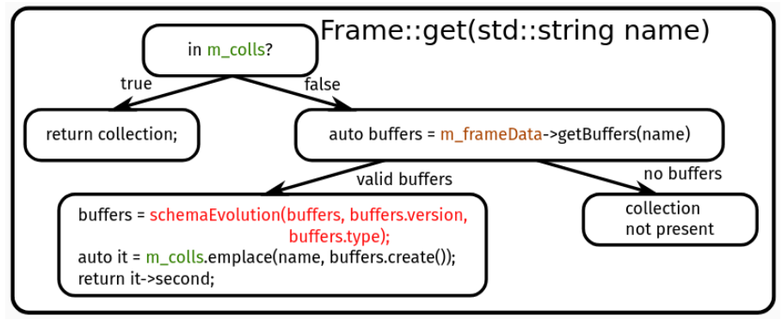
Frames

- Frame is a container aggregating all relevant data
- Defines an interval of validity / category for contained data
  - Event, Run, readout frame, &dots;
- Thread safe interface



Schema evolution

- Introduces visioning to the Collection types
- Handles changes between different versions
- Evolution always to latest version



More details: T. Madlener

FCCANALYSES DATASETS

Plethora of processes are pre-generated and available from EOS

- Two main production campaigns in use:
  - Spring 2021
  - Winter 2023
- Processes are identified by its name, e.g.: `p8_ee_WW_ecm240`
- The production Database browsable at:
  fcc-physics-events.web.cern.ch
- Example:
  Delphes events, IDEA, FCCee, winter 2023
- EOS directory:
  `/eos/experiment/fcc/...`
- Generation handled by EventProducer

PODIO-DUMP

Example:

```
15:32:16 [jsmiesko@fcc-ironic-02 metadata]$ podio-dump -h
usage: podio-dump [-h] [-c CATEGORY] [-e ENTRIES] [-d] [--dump-edm DUMP_EDM] [--version] inpu

Dump contents of a podio file to stdout

positional arguments:
  inputfile             Name of the file to dump content from

options:
  -h, --help            show this help message and exit
  -c CATEGORY, --category CATEGORY
                        Which Frame category to dump
  -e ENTRIES, --entries ENTRIES
                        Which entries to print. A single number, comma separated list of numb
                        "first:last" for an inclusive range of entries. Defaults to the first
```

- Lists collections from a frame
- Can dump all the event details
- Compatible only with `winter2023`

EDM4HEP2JSON

Example:

```
13:28:43 [jsmiesko@fcc-ironic-02 metadata]$ edm4hep2json -h
Usage: edm4hep2json [olenfvh] FILEPATH
  -o/--out-file           output file path
                            default: "?edm4hep.root" --> ".edm4hep.json"
  -l/--coll-list          comma separated list of collections to be converted
  -e/--events             comma separated list of events to be processed
  -n/--nevents            maximal number of events to be processed
  -f/--frame-name         input frame name
                            default: "events"
  -v/--verbose            be more verbose
  -h/--help               show this help message
```

- Dumps all requested collections to a formatted JSON file
- Outputted JSON file can be visualized in Phoenix
- MC Particle tree can be investigated with dmX
- Compatible only with `winter2023`

COLLINFO

Example:

```
13:33:46 [jsmiesko@fcc-ironic-02 Podio (main=)]$ ./collInfo /eos/experiment/fcc/ee/generation/
/winter2023/IDEA/p8_ee_ZZ_ecm240/events_092194859.root
ID   Name                        Type
-------------------------------------------------------------------------
 1   MissingET                   edm4hep::ReconstructedParticleCollection
 2   MCRecoAssociations          edm4hep::ReconstructedParticleCollection
 3   ParticleIDs                 edm4hep::ReconstructedParticleCollection
 4   magFieldBz                  edm4hep::ReconstructedParticleCollection
 5   TrackerHits                 edm4hep::MCRecoParticleAssociationCollection
 6   EFlowTrack                  edm4hep::ParticleIDCollection
 7   CalorimeterHits             podio::UserDataCollection
 8   Particle                    edm4hep::TrackerHitCollection
 9   Photon                      edm4hep::TrackCollection
10   EFlowTrack_L                edm4hep::CalorimeterHitCollection
11   Electron                    edm4hep::MCParticleCollection
```

- Dumps collection ID, name and type in "events" frame
- Lives in FCC AuxTools
- Compatible with `spring2021` and `winter2023`

CONCLUSIONS

- EDM4hep files can be inspected by several tools
  - `podio-dump`, `collInfo`, `edm4hep2json`
- Improvements needed toward visualization of collection relations
- Not all parts of FCCSW stack fully utilize PODIO
  - FCCAnalyses — uses only POD Layer
- EDM4hep will need changes for FCC full simulation
- Policies how to save event data into datamodel in flux

BACKUP

COMPATIBILITY TABLE

`podio-dump`

	stable	nightlies
Spring 2021	x	x
Winter 2023	y	y

`collInfo`

	stable	nightlies
Spring 2021	y	y
Winter 2023	y	y

`edm4hep2json`

	stable	nightlies
Spring 2021	x	x
Winter 2023	x	y

More details: key4hep/EDM4hep #228