ALICE EDM

- **Event Data Model**
  - Formally two aspects
    - Structure and relations
    - Content itself
  - None have to be overestimated (underestimated)
- **Classification by content**
  - Raw data
  - Event Reconstruction Data (ESD)
    - Data Summary Tape
  - Analysis Data (AOD)
    - Mini- Micro- Nano- DST
    - Collaboration wide selections
    - User defined selections
  - Event tag data
  - Simulation Data

ALICE EDM

- **Two aspects**
  - **Event Data Model is just about a collection of bits and bytes**
    - **after all it's DATA what's about**
  - **Structure, pointers, navigation, access methods are essential**
    - **after all we want to read, to store, to handle…**
- **Form versus Content (dialectic)**

Raw Data

- **This is surely more collection of bits and bytes**
  - **Structure has to be find**

- **High level structure imposed by DAQ architecture**
  - **About of 500 DDL - independent data streams**
  - **Each stream will have a standard header per event**
    - **event identification**
    - **sub-event length**
    - **trigger type**
    - **error status**
    - **…**
  - **Then data - bit-byte stream**
    - **coding defined be Front-end Electronics (FEE)**

Raw Data

- **Better write it as soon as possible !**
  - At least at the beginning of experiment
- **What about objectification ?**
  - Maybe, on the high level - per DDL basis
  - But it is not really necessary
    - one can do selection of detector data on DDL identifier as efficient as on object
  - One has to see overhead
- **What about meta-data (it's not my job, I know)**
  - preppended - calibration settings, etc.
  - appended - summary scalers, etc.

ESD (reconstruction)

- **Reduction typically 1/10 of raw data volume**
  - **What we are speaking about is from the point of view of information theory data reduction with minimal information loss**
    - **most of the experiments I participated in on this step in fact enlarge the data volume**

- This structure is defined experiment-collaboration wide
  - It is quite rigid
  - Usually experiment has a standard input ‘routines’ which you get from ‘somebody’
  - Only few people know exactly what are those all numbers are about (covariant matrices, residuals, quality codes etc.)

ESD (reconstruction)

- **Here we have clear structure of objects**
  - **Event contain its header (identification) and**

hasa

primary vertex, which

isa

vertex, which

hasa

primary tracks, which

isa

track

and secondary vertices, which

isa

vertex, which ***hasa*** secondary tracks which ***isa*** track ...

ESD (reconstruction)

What is inside ?

- Everything what can be needed during analysis under any circumstances
- Example: a track
  - its parameters at first measured point (5 numbers)
  - covariant matrix (15 numbers)
  - its parameters at last measured point (5 numbers)
  - covariant matrix (15 numbers)
  - measured length and error (2 numbers)
  - quality parameters (number of points in each detector, mean residuals in each detector, global chi**2, etc. about 10 numbers)
  - particle identification information (about 10 numbers)

  so, we just get about 60 numbers (i.e. about 250 bytes par track)
- Luckily vertices are less consuming
- And methods to get other calculable information
  - OO paradigm gives here large flexibility to hide the physical content

ESD (reconstruction)

- **How we get it ?**
  - Reconstructing !
  - We have incremental track reconstruction model
    - first TPC, then ITS, after TRD, later HMPID etc. (PHOS veto)
  - As I am thinking very similar situation will be also with PID information
  - Therefore track object sniff in more and more information, however its structure is the same
  - Two ways have been discussed
    - let’s sniffed-in fat track object became finally persistent being itself part of EOD (I am a partisan of this)
    - let’s do at each step new object and at the end feed-in a new persistent global track object (which I object, but it is a possibility)
  - This is for a discussion

ESD (reconstruction)

- **Do we need to navigate back to raw data?**
  - **Remember, we said that there is:**

    'Everything what can be needed during analysis under any circumstances'
  - So the simple answer is NO
  - Better say: 'Let's tray to avoid it'
- From the past experience
  - Sometimes (however, very rarely) is needed
  - After all, it is not too difficult to get back just using event identifiers - no special access schemes with pointers
  - It is just not too efficient, but if it is rare who cares ?

Remarks on Raw on ESD

- All the data discussed so far are considered to be strictly once written read-only (or you can put them on CD)
- Consequence: do not invent any data base schemes to store them and protect them (file system is enough)

AODs and TAGs

- **These are more ‘lively’ data**
  - they could ‘select’ and ‘reselect’ and ‘select again’
  - usually a standard set of selection is used collaboration wide which gives ‘collaboration wide AODs’
  - inventive users tend to define their selections ‘user AODs’
- **Both we need to prototype already with simulation**
- **Do we need to navigate back ?**
  - In N-tuple age people usually put inside N-tuple everything which potentially could be needed in order to avoid the pain going back to DST
  - Back navigation can resolve this ‘Mega-N-tuple syndrom’
  - The short answer is ‘I don’t know’
- **TAGs - optimize the selection performance**

Discussion

- **We have to discussed The Event Data Model**
  - **see for example ATLAS discussion on**

  http://documents.cern.ch/AGE/current/fullAgenda.php?ida=a01922#s6?ida=a01922
- **But we have to start to play with (or prototype ?)**

Event Summary Data

- **Event identification**
  - event number
  - run number
  - trigger type bitmap
  - version of reconstruction
- **Summary information**
  - bitmap of Data Detector Link participate
  - number of secondary vertices
  - N participants estimate
  - N participants error
  - number of electrons
  - number of muons
  - number of pions
  - number of kaons
  - number of protons
  - number of photons in PHOS
  - number of charged clusters in PHOS
  - number of neutrons in PHOS
  - number of clusters EMCAL
  - number of clusters in PMD
  - transverse energy of biggest cluster
  - biggest transverse momentum of charged particles
  - charged multiplicity (few numbers)
  - total transverse energy

Event Summary Data

- **Primary vertex object**
  - number of primary tracks
  - coordinates (3)
  - error matrix (6)
  - container of primary track objects
  - effective mass
  - and its error
- **Container of non-assigned track**
- **Container of secondary vertex objects**
- **Container of photon object**
- **Container of neutrons**
- **Container of EMCAL clusters**
- **Container of PMD clusters**

Track summary object

- Track id
- Track parameters at vertex
  - 5 param
  - 15 matrix elements
- Track parameters at first measured point (w/o vertex constraint)
  - 1+5 param
  - 15 matrix elements
- Track parameters at last measured point
  - 1+5 param
  - 15 matrix elements
- Track length + error
- Index of stop vertex

- Number of assigned points in ITS
- Number of points in TPC
- Number of points in TRD
- Average residual in ITS
- Average residual in TPC
- Average residual in TRD
- Global chi**2
- Particle type
- PID probability for pi
- PID probability for K
- PID probability for p
- PID probability for e