Metadata: Point of View of Physics

Borut Paul Kerševan, Junji Toyo

What is Metadata Physicists Want?

- The answer is **everything** related to a sample:
  - Statistics, visible cross-section (cross-section X gen. Efficiency), pcaches used, settings in event generation/G4/digi/reconstruction...
  - Good organization and presentation of this information is of the essence:
    - Links to more information on settings/parameters used.
    - Easy search and navigation.
    - 'Nice'/user friendly presentation.
  - An important point is the workload it presents – as automated as possible..
  - The policy/aim is to have all the Metadata information available centrally in AMI.
    - Groups still use Twiki pages to track their production requests: We need to understand why!
    - It would be good if as little 'manual info as possible' would be present at the very least and as much as possible through AMI.

Recent Improvements..

- Links to twiki pages were added: e.g. for Geometry etc..
  - Some twiki links missing (easily added I think), like trigger configurations etc.. (we can work out a list).
- Automation: improvements for EvGen: the cross-section and gen. filtering eff. are now processed fully from jobOptions; and combined to give accurate predictions.
  - An ongoing effort within MC group to further improve on this for all MC generators.
- Possibility to extract bookmarks.
- Command line tools development.

Presentation

- A lot of information is available but the user interface somewhat 'technical'.
- Information organization sometimes organized according to production, not physics logic.
  - Example: somewhat difficult to get the sample cross-section:
    - One usually looks for AODs, then needs to use the 'provenance' since the cross-section info is only visible in the EVNT sample where it was picked.. (most common question on mail threads).
- Input from physics users on the design improvements?

Element's information

logicalDatasetName	mc09_7TeV.115257.PythiaUED_6_1000.merge.AOD.e590_s765_s767_r1427_r1429 DQ2 - GANGA export - Provenance - Series
dataType	AOD
physicsCategory	
physicsSubCategory	
physicistResponsible	
nFiles	2
totalEvents	9996
physicsComment	
physicsProcess	
TransformationPackage	15.6.12.1
AtlasRelease	15.6.12
prodsysStatus	EVENTS_AVAILABLE
version	e590_s765_s767_r1427_r1429 Config_Tag - Datasets
datasetNumber	115257
jobConfig	
principalPhysicsGroup	phys-exotics
physicsShort	PythiaUED_6_1000
requestedBy	borut.kersevan@ijs.si
creationComment	MC09 production - EXOTICS

Children elements

prodsys_task	1 Records
dataset_extra	14 Records
dataset_keywords	No records found
dataset_comment	No records found
files	2 Records
jobOptions	No records found

field	fileType
value	aod

field	beamType
value	collisions

field	--athenaopts
value	NONE

field	--ignoreerrors
value	NONE

field	--omitvalidation
value	NONE

field	DBRelease TWIKI
value	11.6.1

field	autoConfiguration

Navigation

- Contextual searches: One would like e.g. to find samples with GEO-10-08-00, done in May reprocessing by Egamma group and Pythia generator.
  - **The two example cases will trigger complaints to me if people can't find the samples they requested:**
    - Some inconsistencies like the physics group info propagation (e.g. on above example, I know the request exists but..)
    - Pileup samples mis-identified (this goes also for e.g. ProdDash..).
  - **Reprocessing/campaign info not present:**
    - We need to sort this out; it would be good to have a common place where prod. tags can be put in logical groupings and are accessible by AMI, ProDsys, DDM … (we started the discussion on this in ADC). Currently only a couple of wikis and mail threads..

Other Considerations

- Conflicting information of Prodsys vs AMI vs DDM is a major source of confusion when it happens!
  - This should really be avoided if possible. Improve synchronization?
    - Unclear messages really need to be avoided: 'EVENTS AVAILABLE/DELETION STARTED' encountered a couple of times leading to a lot of panic..
  - Personally, I don't find metadata info being in several places a problem, if it helps:
    - e.g. I would love to be able to get the total AOD size for e/gamma group for May reprocessing by running a DDM command..
- What should be worked upon is the correct procedure for feature/development requests:
  - It happened that something I asked to be changed/added in ProdSys caused problems in AMI (e.g. new NTUP names, configuration changes..)