# Analysis Model Forum 13-14 September @ CERN Summary



- Goals for the meeting
- DPD Content: Physics analysis use cases
- DPD Content: Performance group issues
- Implementation issues
- Tools and strategies
- Final remarks

# Goals for the meeting

Ian Hinchliffe

- Develop and understand DPD use cases
- What tools are needed when working in ROOT
- DPD: How many and how often?
- Baseline:
  - CBNTAA and SAN not anymore produced by default in rel 13
  - AOD can be analyzed directly from ROOT
  - DPD could be a skimmed/slimmed/thinned AOD
  - Need to converge on decision by end of the year

# DPD content: Physics analysis use cases

- Questions to be answered:
  - How many different contents were used? 1 per group, 1 per analysis? 1 per CSC note?
  - How much ”user data” and derived quantities were added to the content obtained directly from AOD?
  - What fraction of the information was skimmed, thinned or slimmed away?
  - How many times did you reprocess the entire data set?
  - After re-processing was it necessary to retain the output from earlier passes?

# DPD content: Physics analysis use cases ► Response

T. LeCompte, S. Paganis, A. Shibata, J. Catmore, K. Black, P. De Jong

- number of DPDs varies a lot for different physics groups
  - from basically one common format (`TopView`) in the top group over one to a few per analysis (Higgs) to individual DPDs for each group member (Exotics)
  - most groups agreed that it would be possible to have a common loose DPD for the group from which stricter analysis bound DPDs can be derived
- skimming has not been used much
  - mostly due to nature of simulated data (filename already tells you the process, generator cuts, etc.)
  - will be used on real data

# DPD content: Physics analysis use cases ► Response cont'd.

T. LeCompte, S. Paganis, A. Shibata, J. Catmore, K. Black, P. De Jong

- slimming and thinning patterns varied again
  - fear to get it wrong and throw useful info away is high
  - ► best adopted at the stricter DPD not so much at the general group DPD
- earlier DPD versions deleted (or planed to do so) as soon as replacements were validated
  - ► for some period twice the entire DPD set has to be kept

# DPD content: Performance group issues

- Questions to be answered:
  - What can be done on AOD/DPD?
  - How often is it needed to access ESD/RDO to re-make AOD/DPD?
  - What are the computing resources needed?
  - Can some calibrations/re-reconstructions be performed on AOD/DPD in a maybe coarser way instead of going back to ESD/RDO?
  - How can the conditions data version applied be identified on each level?

# DPD content: Performance groups issues ► Response

S. Haywood, D. Froidevaux, P.-A. Delsart, O. Kortner

- AOD/DPD operations for performance groups mostly require athena
  - vertexing and b-tagging needs access to B-field, geometry etc.
  - e/gamma re-calibration has to use proper tools and conditions data like ID material
  - cluster based MET and jets can be re-calculated on AOD/DPD also outside athena except for calibration
  - muon re-fitting, calibration and alignment possible on AOD inside athena; some muon performance checks possible on DPD
  - ► AOD access by performance groups mainly within athena; DPD of limited use
- RDO/ESD access pattern
  - ESD sufficient for almost all performance group tasks
  - ESD needed for some important tasks
  - ► access to RDO not in general required but ESD access vital for most performance groups for validation, systematics, efficiencies, etc.

# DPD content: Performance groups issues ► Response cont'd.

S. Haywood, D. Froidevaux, P.-A. Delsart, O. Kortner

- computing resources needed
  - in general computing time less an issue than access to ESD (storage)
  - ► need prioritized access to ESD for performance groups
- coarse AOD/DPD-based calibrations/re-reconstructions
  - not for e/gamma, tracking and muons
  - maybe to some degree for jets and MET
  - ► need to write and test tools to explore this for jets and MET
- conditions data version not stored in reco objects
  - ► rely on event header and provide user friendly interface to access conditions data version

# Implementation issues

S. Snyder, W. Lavrison, S. M., J. Komaragiri, S. Paganis, R. D. Schaffer

- `AthenaROOTAccess`
  - Overview
    - AOD/DPD access as `TTree` from ROOT-session
    - Python, CINT, or C++ access possible
    - no athena framework
  - need to maintain functionality for future releases and develop a smaller release kit for ROOT based AOD/DPD access
  - Performance issues
    - athena and ROOT use same IO
    - .so loading and large dictionaries huge time consumers for both
  - work on dictionaries and ROOT benefit both
  - Tutorial
    - skeleton examples in `AthenaROOTAccessExamples`
    - easy to extend
    - CINT considerably slower than C++
  - need corresponding python examples to evaluate performance
  - identify list of athena tools that can be used with `AthenaROOTAccess` – e.g. can one run EventView's overlap removal?

# Implementation issues, cont'd.

P. Faccioli

- Interactive athena experience
  - full athena framework plus python prompt
  - access to StoreGate containers via PyTools
  - can this be extended by a CINT prompt?

# Tools and strategies

F. Winklmeier, S. Binet, T. Lari

- DPD strategies from BaBar
  - centralized production of DPDs (Skim production)
  - validated set of common filters and tools for skimming and user data
  - looks promising but need to translate this to ATLAS use case
- Skimming/Thinning/Slimming Tools
  - Skimming: keep interesting events only
  - Thinning: keep interesting objects only
  - Slimming: keep interesting properties only
  - Thinning probably most critical; feedback needed
- DPD making tools ouside `EventView`
  - lightweight `SusyPlot` to select objects, remove overlaps and write to DPD (ntuple in this case)
  - similar `UserAnalysis`-based examples to create ntuples from AOD are in use
  - or just writing out a thinned AOD
  - framework independent tools are important

# Tools and strategies, cont'd.

A. Farbin, R. Jones

- User data and `EventView` for DPD making
  - CSC style access/analysis patterns don't scale to LHC data volumes
  - `EventView` provides framework and common tools
  - thinning/slimming with `HighPtView`
  - different performance and phsyics DPDs
  - soft overlap removal with `ParticleView`
  - persistifying `EventView` needs work
  - need to ensure that it is useable early in a release cycle to be adopted by users
  - decouple tools from framework
- Computing model issues
  - DPD a must because of speed/size/portability
  - full group/stream DPDs on Tier 3s problematic (disk-space/bandwidth)
  - Tier 2s still low on disk
  - scheduled group analysis at Tier 1s (in trains?)
  - need more experience with TAGs
  - need to organize on-demand analysis

# Final remarks

- Future meetings
  - Two day meeting 28-29 November: Fixed.
  - Decision on DPD will be made after this meeting
  - Proposed two half day afternoon phone meetings before then
  - October 15 or 16?
  - November 7 or 8?