nanoAODplus: news

Nano meeting

DESY, 12.3.2020

Achim Geiser, DESY Hamburg
For the "nano" team

- Continuation of work started in summer 2018,
  many thanks to all previous participants!
- 12 months of EPR allocated for 2020, like 2019 (had asked for 15)
- Negotiations concerning DESY institutional responsibility ongoing
- Pledge part of the work (beyond 12 months) through different categories?

For details about tasks and concept see backup and previous contributions to this meeting

Tentative list of tasks/contents for 2020

- nanoAOD ntuple content (9_4 v2): fully done , partially done, already useful/used for analysis, being implemented , other

	variables	implemented	content implemented	content validated	remaining work
run/event/lumis.	3	100%	100%	100%	done
Generator /PSweight	11	-	-	-	~0.5 months, Melanie? (Hannes??)
PV /OtherPV /Pileup	14	70%	70%	50%	~0.25 months, Achim half done
SV	13	100%	10%	-	~0.5+0.5 months, Josry, Achim
GenPart	9	100%	50%	20%	~0.25 months, Achim advanced
Muon	35	100%	80%	75%	~0.5 months, Achim almost done
Electron	48	55%	50%	45%	~2.5(+0.5) months, Melanie (+Nuha) half done
Photon	28	25%	25%	-	~2 months, !not yet covered! (use H->gg?)
Tau	38	25%	25%	-	~3 months, !not yet covered! (use H->tau?)
IsoTrack	13	100%	-	-	~0.5 months, Achim
GenDressedLepton + GenVisTau	14	-	-	-	~0.5 months, N.N.
Jet+FatJet +SubJet +SoftActivityJet	79	10%	10%	-	~3.5 months, Armando? Heng? Josry? AG?
GenJet +GenJetAk8	14	-	-	-	~0.5 months, N.N.
MET+TkMET +CaloMET +RawMET+PuppiMET	23	30%	30%	-	~0.5 months , N.N.
TrigObj advanced	11	55%	45%	20%	~0.5+0.5 month, Achim, Qun?
HLT	569	100%	100%	100%	done
LHEPart +LHEPdfWeight + LHEScaleWeight+LHEWeight_originalXWGTUP	11	-	-	-	~0.5 months Josry ? (Hannes??)
Flag	26	100%	-	-	~0.5 months, !not yet covered!
Various other	10	-	-	-	~1 month , !not yet covered!

- implementation of nanoAOD header — ~0.5 months, Hamed?
- nanoAOD tools — ~0.5 months , Nuha
- coordination + set up & manage twiki + git repository — ~1 month, Achim
- various general technicalities of setup — ~0.5 months, Afiq

total — ~26 months 2019/20

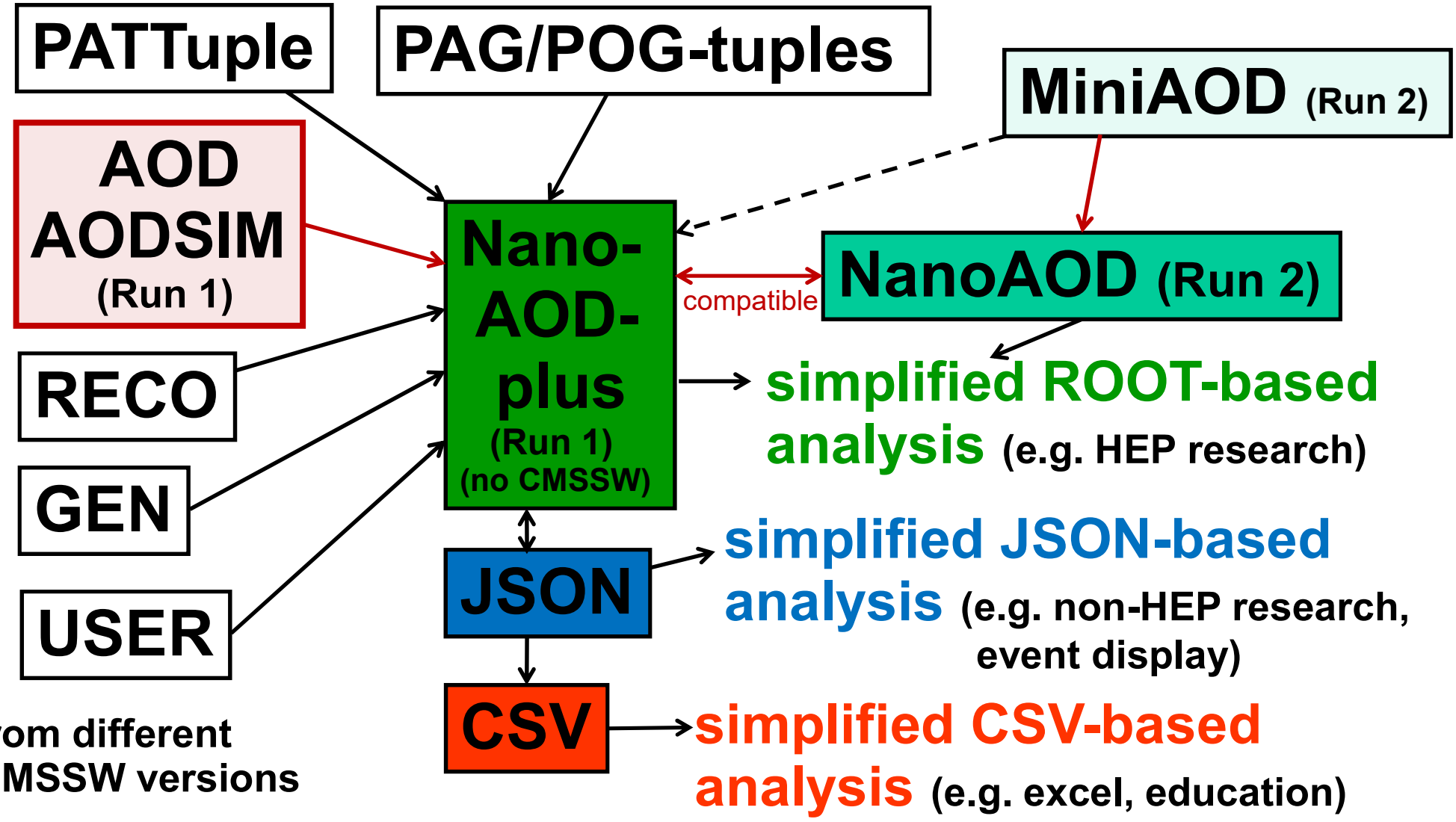
~15 this year

~6.5 not covered/next year (tbc)

Backup

Thoughts about simplified DPOA data format: CMS

Design common flat ntuple format for all datasets (remove CMSSW dependence)



Motivation/goals for nanoAODplus format for Run 1

- **independence from `old' CMSSW versions** (or CMSSW in general)

  -> analysis in non-CMS environment,

  no need for virtual machines or container encapsulation

- **CMS members can run Run 2 nanoAOD-based analyses also on Run 1 legacy data and vice versa with same code**
  (also outsiders once Run 2 data will be released as Open Data)

  -> **identical** nanoAOD variable **names**

  -> **same variable content** **(as much as possible)**

  -> task:

  recode Run 2 algorithms for nanoAOD content directly from basic AOD variables, such that they work for CMSSW 4_2_8, 5_3_32 (Run 1 legacy), as well as 7_X (2015, no nanoAOD so far) and 8_X/9_X/10_X (for cross-validation with official Run 2 nanoAOD)

- **Add specific Run 1 variables ("plus")**

Technical implementation of ntuple production

- **EDanalyzer (NanoAnalyzer) which compiles/runs**
  in VMs or containers (for SL5)
  on DESY Tier 2 farm (for SL6/SL7)
  or with CRAB (via containers for SL5)

  for technicalities
  see presentation N. Jomhari
  at DPOA meeting March 13

  (single code, different configurations, differences between CMSSW versions accounted for via #ifdef flags)

- **Input is AOD (working on miniAOD interface for debugging)**

- **Implement Run 2 nanoAOD algorithms (according to workbook)**
  on Run 1 AOD whenever technically possible

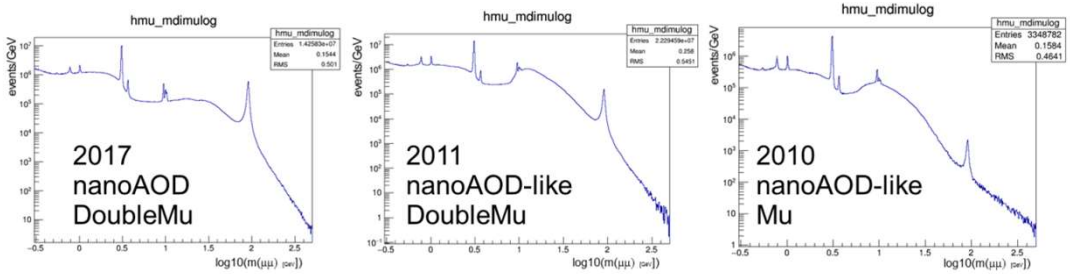
- **In addition, implement legacy Run 1 algorithms (extra variables, according to legacy workbooks) whenever useful (plus some further variables)**

- **Output is flat Root ntuple with nanoAOD variables, currently accessible on DESY dcache via XRootD (working on DBS publication option)**

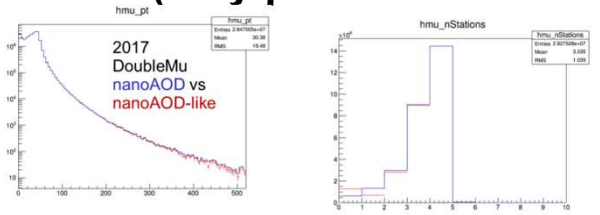
- **Twiki Documentation (under development) :**
  https://twiki.cern.ch/twiki/bin/viewauth/CMS/DPOANanoAODlike

Validation tools and strategy

- **Indirectly compare some physics distributions for different datasets**
  examples see presentation at fall C&O meeting:



- **Directly compare technical distributions (only possible for Run 2)**
  examples see presentation at fall C&O meeting:



- **New: Use BuildIndex and Friend functions of Root to compare nanoAOD and nanoAOD-like variables event-by-event, even if input event sets only partially overlap and events occur in different order (only possible for Run 2)**
  (thanks to A. Ricci and J. Metwally for support!)

  -> can validate and debug exactly

- **Exactly reproduce some known/well-validated Run 1 distributions from nanoAOD-like ntuple**

Conclusions and Outlook

nanoAOD-like data format for Run 1 making progress,
now organised through dedicated DPOA tasks
-> strengthen interaction with XPOG, POGs/PAGs, and PPD

tasks defined (see backup) and person power (EPR) for this year being assigned
(team of ~10 people part time) ***** today's meeting! *****

-> hope to complete nanoAODplus ntuple for Run 1 by end of 2020,
in parallel to Run 2 super-legacy processings

-> all legacy data should be analysable in nanoAOD(like) format
with the same CMSSW-independent Root analysis code,
and (as much as possible) with the same variable content

eventually available as Open Data together with AOD/miniAOD
-> easier for outsiders to do analysis compared to current Run 1 AOD

plans

nanoAOD-like data format for Run 1 making progress,

first actual applications in sight

-> hope to complete for Run 1 within next two years,
in parallel to Run 2 super-legacy processings
-> all legacy data should be analysable in nanoAOD(like) format

current situation:

Table:	0.9	2.76	5	7	8	13 TeV
ep	ZEUS*					
pp	2010/17	2010/13	2015/17	2010/11	2012	2015
						2016/17/18
pPb			2012/13/15		2016	
PbPb		2010/11	2015			

*external project in preparation

AOD
nanoAOD available
miniAOD available
RECO only