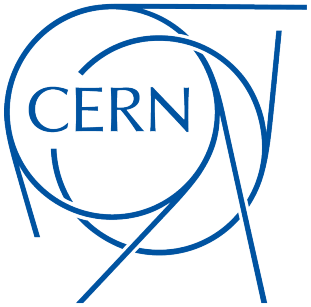

Proposal for the new naming scheme of the MC production cycles in ALICE

Igor Lakomov
Alexander Kalweit
Evgeny Kryshen
with input from CERN-ALICE group

Offline Forum, 18/03/2015

Motivation and constraints

Motivation

- To allow user to find fast and easily a relevant MC production to be used for the analysis:
  - ✓ to make clear names of the MC productions so that the user knows almost immediately the general information about the MC production.
  - ✓ to create a reliable data base allowing the more detailed search

Constraints

✧ The name should not be too long.
(now the longest name is “LHC12a17b_fix_jemalloc” – 22 symbols)

✧ Should be easy to understand at least **to which data it is anchored** and the responsible PWG.

✧ One should know what is the latest version of the same MC productions.

New naming scheme proposal

The name is divided into three parts:

- ✓ The first two parts are given in a pre-defined way.
- ✓ The last part is defined by user.
- ✓ +in the end of the name we put a version number of the MC production

- **First** part is related to the period and pass number the MC production was **anchored** to.
- **Second** part defines the **PWG** responsible for this MC production (DQ, UD and etc.)
- **Third** part is defined by **user** in a fixed length (~10 symbols – to be discussed)

- See example on the next page

New scheme proposal: example

MC10de_p2_UD_cga_jps_en_1
MC10de_p2_UD_cga_jps_mn_1
MC10de_p2_UD_cgg_gag_en_1
MC10de_p2_UD_cgg_gag_mn_1
MC10de_p2_UD_cgp_jps_en_1
MC10de_p2_UD_cgp_jps_mn_1
MC10de_p2_UD_cgp_psi_el_1
MC10de_p2_UD_cgp_psi_en_1
MC10de_p2_UD_cgp_psi_et_1
MC10de_p2_UD_cgp_psi_ml_1
MC10de_p2_UD_cgp_psi_mn_1
MC10de_p2_UD_cgp_psi_mt_1
MC10de_p2_UD_sga_jps_mn_1
MC10de_p2_UD_sgg_gag_mn_1
MC10de_p2_UD_sgp_jps_mn_1
MC10de_p2_UD_sgp_psi_ml_1
MC10de_p2_UD_sgp_psi_mn_1
MC10de_p2_UD_sgp_psi_mt_1

First part:
MC10de_p2 – anchored to LHC10d and LHC10e pass2

Second part:
UD – corresponds to the PWG-UD

Last part*:
c,s – in central/semi-forward rapidity
gg,gp,gA – gamma-gamma/gamma-p/gamma-nucleus interaction
jps,psi,gag – jpsi/psi(2S)/gamma-gamma production
e,m – dielectron/dimuon decay channel
n,t,l – neutral/transverse/longitudinal polarization

1 – first version of this MC

*Here the user part consists of 10 symbols just as example but it can be fixed to a lower value.

To be discussed

- What is the most important information to be added to the name of the MC production?
- To allow a backward compatibility, i.e. matching between new naming scheme and the existing MC productions, previously produced: create “aliases”?
- What to do in **exceptional** cases?
  1) few datasets used as anchored:
     a) if they are all from one year -> to add corresponding period letters (i.e. MC10de_p2 for pass 2 of LHC10d and LHC10e)
     b) if more complicated (as for the LHC14i2) -> to create separate folders for each data period used
  3) few PWG involved -> to name «MX» instead of the PWG name and/or to add names of the corresponding PWGs; put «MB» if for all the PWGs
- To update the Config.C for the automatic naming of the new MC production (create new fields (variables) which are filled by the MC creator: PWG, anchored data sets, comment, user defined part of the name, version).
- To think of the corresponding fields in JIRA for faster searching.
- Some rules should be agreed on the comments to all MC productions to standardize them, i.e. the order of the given information and the pattern for each piece of information.

To be discussed

The last free bullets can be also used for the new data base creation.

(see Marian's talk)

- To update the Config.C for the automatic naming of the new MC production (create new fields (variables) which are filled by the MC creator: PWG, anchored data sets, comment, user defined part of the name, version).
- To think of the corresponding fields in JIRA for faster searching.
- Some rules should be agreed on the comments to all MC productions to standardize them, i.e. the order of the given information and the pattern for each piece of information.

Conclusion

- Something should be done to ease the life of analyzers, especially students and newcomers.
- Two steps are proposed:
  1. Fast: Change the MC naming scheme and standardize the comment to the MC.
  2. A bit longer but with additional advantages: Create a new DB.
- Both of them help analyzers and complement each other.
- Both of them can be discussed and be changed a bit.
- For the DB proposal some feedback is needed from the PWG conveners and Offline experts.
- Not easy to automatically create the DB without the standard rules for the comment part.

Backup

Alternative scheme proposal

The full names are given in a pre-defined way.

Example: MC10d_p2_000123_v01 is anchored to LHC10d pass2

digicode 000123 defines the MC options:

0 – extra options (if needed)
0 – extra options (if needed)
0 – extra options (if needed)
1 – PWG responsible for the production
2 – generator used (pythia/perugia….)
3 – system of detectors (central barrel/muon arm…)

v01 corresponds to the first version of this MC production

- First part is related to the period and pass number the MC production was anchored to.

- Second part is a digicode where every digit corresponds to some MC options.
(this digicode is analogous to AliBits usage)

- +version number in the end of the name

Fields in JDL useful for the naming

```
User = "aliprod";
 JobTag = "comment: Production of muons from W/Z with POWHEG anchor LHC13d,e,f - ZPOWHEG_pp_pPb_LHC13e_LHC13MisAlignCDB_5 -
ALIROOT-5632";
 Packages = {
       "VO_ALICE@AliRoot::v5-05-Rev-24",
       "VO_ALICE@GEANT3::v1-15a-1",
       "VO_ALICE@ROOT::v5-34-08-6",
       "VO_ALICE@APISCONFIG::V1.1x"
};
 JDLPath = "/alice/cern.ch/user/a/aliprod/LHC14l2e9/JDL”;



 LPMProductionTag = "LHC14l2e9";
 LPMProductionType = "MC";
 LPMRunNumber = "196310";
 LPMAnchorProduction = "LHC13e";
 LPMAnchorRun = "196310";
 LPMAnchorYear = "2013";
 LPMMetaData = "Packages=[VO_ALICE@AliRoot::v5-05-Rev-24, VO_ALICE@GEANT3::v1-15a-1, VO_ALICE@ROOT::v5-34-08-6,
VO_ALICE@APISCONFIG::V1.1x];OutputDir=/alice/sim/2014/LHC14l2e9/196310/
#alien_counter_03i#;LPMRunNumber=196310;LPMProductionType=MC;LPMProductionTag=LHC14l2e9;LPMAnchorRun=196310;LPMAnchorPr
oduction=LHC13e;LPMAnchorYear=2013";
 PWG = "COMMON";
 InputDataType = "NONE";
 Activity = "SIM+AOD”;
```

Advantages of the new DB

- Fast search of the MC production (user can use filters in different fields).
- Allows sorting by names and other fields.
- Can be automatically updated based on the fields in Config.C and/or fields in JIRA.
- Can give a matching between the data and MC in the table.
- Can provide information about the QA status for detectors and PWG and etc.
- Scaling: new fields can be added if needed.
- Additional info: https://indico.cern.ch/event/373581/contribution/0/material/slides/4.pdf

MC productions statistics

From Marian

MC productions:

- 319 productions registered at https://alimonitor.cern.ch/PWG/
  - production table available until July 2013
- 692 productions registered in alien production folder (15.02.205) /alice/cern.ch/user/a/aliprod
- 164 production with QA automatically published on the central QA web page
  - automatic QA since 2012
  - automatic QA web page creation since June 2014 (TPC,TOF)
    - some detector joined later therefore amount of the production smaller (TRD 50)

O(10^3) productions generated in RUN1

The same order of magnitude can be expected in RUN2, RUN3

Some DB support needed in order to handle information and to make automatic procedures possible

Do we need the automatic procedures ?