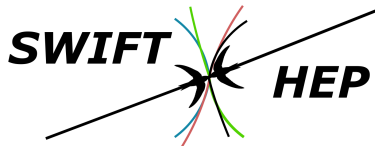


# MC production and validation

Christian Gütschow

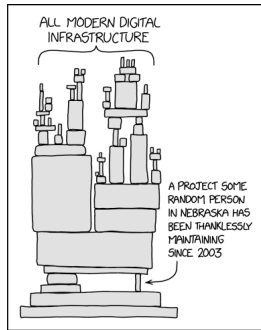
MCnet + KEY4HEP meeting

05 December 2024

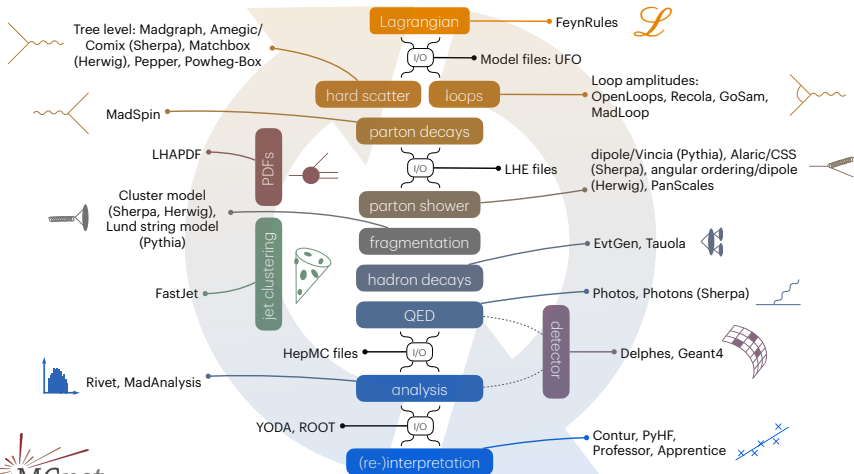


## Overview

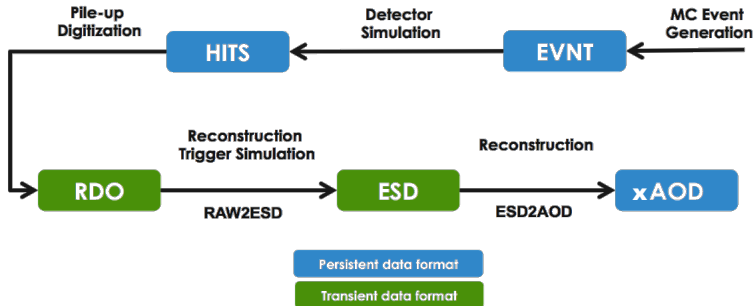
- ATLAS MC production system is a complex infrastructure to orchestrate generation, simulation, reconstruction, and storage of MC datasets as well as metadata handling
- Sherpa introduced a fully automated workflow for physics validation this year



[xkcd.com]

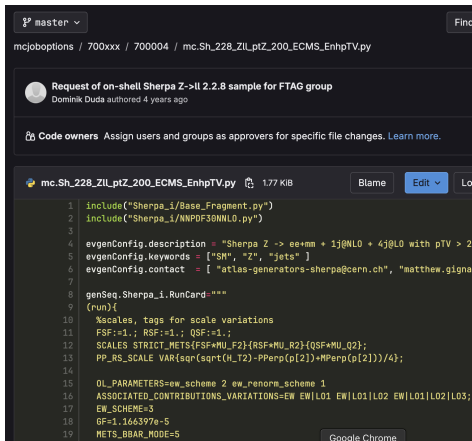


## Typical ATLAS MC production workflow



- external tools installed for different architectures on `cvmf`s in [LCG releases]
- set of generator releases collected into a tagged Athena release for central production
- all jobs run on the WLCG, output files and log accessible through [rucio]

## MC generation job steering



The screenshot shows a GitHub repository page for the file `mc.Sh_228_ZIL_ptZ_200_ECMS_EnhptV.py`. The file content is as follows:

```

1 include("Sherpa_i/Base_Fragment.py")
2 include("Sherpa_i/NNPDF30NNLO.py")
3
4 evgenConfig.description = "Sherpa Z -> ee+mm + 1j@NLO + 4j@LO with pTV > 2
5 evgenConfig.keywords = ["SM", "Z", "jets" ]
6 evgenConfig.contact = [ "atlas-generators-sherpa@cern.ch", "matthew.gign
7
8 genSeq.Sherpa_i.RunCard=""
9 (run){
10 %scales, tags for scale variations
11 FSF=1.; RSF=1.; QSF=1.;
12 SCALES_STRICT_METS{FSF*MU_F2}{RSF*MU_R2}{QSF*MU_Q2};
13 PP_RS_SCALE VAR{sqrt(sqrt(H_T2)-PPerp(p[2])+MPerp(p[2]))/4};
14
15 OL_PARAMETERS=ew_scheme 2 ew_renorm_scheme 1
16 ASSOCIATED_CONTRIBUTIONS_VARIATIONS=EW EW|L01 EW|L01|L02 EW|L01|L02|L03;
17 EW_SCHEME=3
18 GF=1.166397e-5
19 METS_BBAR_MODE=5

```

→ Generator run cards wrapped in Python steering snippets (“JobOptions”) for Athena

→ cron job pulls [MCJobOptions repo] onto cvmf.s

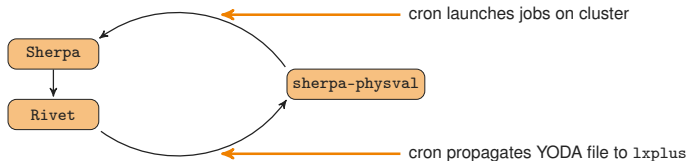
## Metadata handling and dataset discovery

The screenshot shows the ATLAS metadata browser interface. At the top, there are navigation tabs for 'Datasets', 'Files', 'SW Images', 'AMI-Tags', 'Nomenclature', 'Tools', and 'Issue reporting'. Below this is a search bar and a list of dataset categories. The main content area shows a search for 'mc15' with a dropdown menu set to 'DATASET'. A table of results is displayed with columns for 'LOGICALDATASETNAME', 'PRODSYSSTATUS', 'DATATYPE', 'VERSION', and 'NFILES'. The first row is highlighted, showing a dataset with 'ALL EVENTS AVAILABLE' status.

Metadata	
LOGICALDATASETNAME	mc15_13TeV.700004.Sh_228_Z0_pZ_200_ECMS_EnhpTV.wgen.EVNT.e794... #hashtag - BigPanda - Bc1a - Prominence - Synchronise - Series
PRODSYSSTATUS	ALL EVENTS AVAILABLE
DATATYPE	EVENT
VERSION	e7947 Dataset: ASB-Tags
NFILES	1250 Files
TOTALEVENTS	12500000
CROSSSECTION	6.13 pb
GENFILTERF	1
TOTALSIZE	315 GB
ECMENERGY	13.0 TeV
COMPLETION	100 %
STATSALGORITHM	exclude_outliers
PROJECTNAME	mc15_13TeV
PHYSICSCOMMENT	Sherpa Z -> ee+mm - 1j@NLO + 4j@LO with pTV > 250 GeV.
PHYSICSSHORT	Sh_228_Z0_pZ_200_ECMS_EnhpTV

- metadata collected in [central database]
- accessible through the browser as well as via the command line
- critical for both dataset exploration and generator cross-section extraction

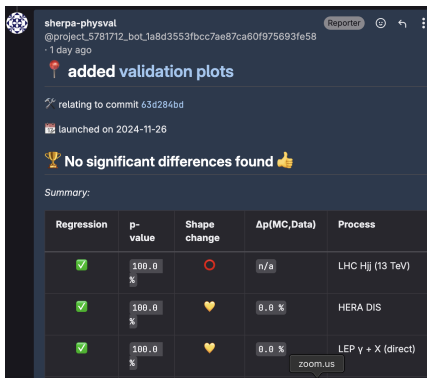
## Automated physics validation in Sherpa



- Based on off-the-shelf Sherpa3+Rivet4 and  $\mathcal{O}(200)$  lines of Python
  - Cron jobs set up at HPCs at UCL and Göttingen to do the heavy lifting
- Check once or twice a day whether the **latest commit message** on any of the feature branches contains a trigger
  - Example: `git commit -m"meaningful message [physval]"`
  - or specify appropriate subgroup(s): `[physval=lep,zinc]`

## Plotting and summary reports

- Cron job runs every 4h to check if there are new YODA files
- runs `rivet-mkhtml` on every new YODA file for the [[validation web page](#)]
- an automated comment is send to the MR page with a link to the validation plots



The screenshot shows a GitHub comment from user 'sherpa-physval' on a pull request. The comment title is 'added validation plots' and it includes a link to the validation plots. The comment text states 'No significant differences found' with a thumbs up emoji. Below the text is a summary table with the following data:

Regression	p-value	Shape change	$\Delta p(\text{MC, Data})$	Process
✓	100.0 %	○	n/a	LHC Hjj (13 TeV)
✓	100.0 %	♥	0.0 %	HERA DIS
✓	100.0 %	♥	0.0 %	LEP $\gamma + X$ (direct)