Analysis preservation in the heavy-ion world

Przemek Karczmarczyk
for the ALICE Collaboration

przemyslaw.karczmarczyk@cern.ch

ISMD2021, 14.07.2021

Overview

Why preservation is essential?

- Preservation in particle physics experiments is essential:
  - so that most of the scientific knowledge obtained can be conserved
  - so that the investments in high-energy physics experiments are not lost
  - for future generations of scientists to keep, reuse, and build upon
- Provides many benefits:
  - Perform old analysis on new MC data
  - Post-publication reproducibility of the MC results for resolving conflicts between published data/MC comparisons

Who can benefit from that?

- Experimental collaborations:
  - Facilitate cooperation between physicists and teams within a given collaboration
  - Simplify the knowledge transfer between scientists
  - Provide a way to re-perform analysis of the old MC data
- HEP community:
  - Apply new theoretical models to the older analysis
  - Resolve conflicts between published results
- Outside HEP community:
  - Provide a way for young students to get acquainted with scientific work by performing a sample analysis
  - Enable the engagement of external communities

What do we want to preserve?

- All the necessary components needed for reusing an analysis and/or dataset to go from MC to the analysis result, such as:
  - MC analysis software
  - documentation
  - processed data
- This will provide all the benefits of preservation we need on the MC front
- For the preservation of experimental data with accompanying software and documentation see OpenData project: http://opendata.cern.ch/

How do we want to do that?

- How to make preservation possible?
  1. Develop tools that can be used for the analysis preservation
  2. Make these tools universal so that anyone within the HEP community can use them for their analysis
  3. Implement policies among collaborations allowing/enforcing them to follow the rules on what components to preserve and how to do that

How do we want to do that?

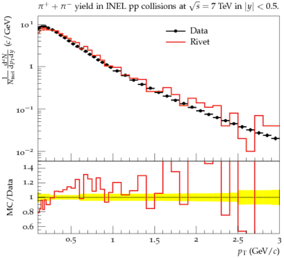
- How to make preservation possible?
  1. Develop tools that can be used for the analysis preservation
  2. Make these tools universal so that anyone within the HEP community can use them for their analysis
  3. Implement policies among collaborations allowing/enforcing them to follow the rules on what components to preserve and how to do that

Preservation tools

- What do we need:
  - A tool that will enable to store scientific results in a central repository
    - Solution: HepData
  - A tool that allows to store MC analysis algorithms in a place accessible to anyone
    - Solution: Rivet
  - A tool that will serve as an interface between generator and analysis
    - Solution: HepMC
  - Other tools that can assist with analysis preservation effort
    - OpenData and other tools

Rivet

- Rivet: Robust Independent Validation of Experiment and Theory
  - Generator-independent analysis framework, arXiv:1003.0694
  - System for validation and tuning of event generators
  - Data & analysis algorithms preservation
- Rivet features:
  - Read input from HepMC format (file, FIFO, or direct in memory)
  - Provides a large set of experimental analyses - currently O(1000)
  - Run one (or more) analyses on the input data from generator
  - Fast and direct comparison between exp. data and generators



Example of Rivet results

Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	e^+e^- (≥ 12 GeV)	e^+e^- (≤ 12 GeV)	Tevatron	RHIC	SPS	Other
Rivet wanted (total):	259	320	427	246	16	503	715	536	1131	454	62	1
Rivet REALLY wanted:	36	42	83	8	0	13	1	0	5	1	0	0
Rivet provided:	26/285 = 9%	175/495 = 35%	91/518 = 18%	16/262 = 6%	8/24 = 33%	9/512 = 2%	180/895 = 20%	305/841 = 36%	58/1189 = 5%	8/462 = 2%	4/66 = 6%	112/113 = 99%

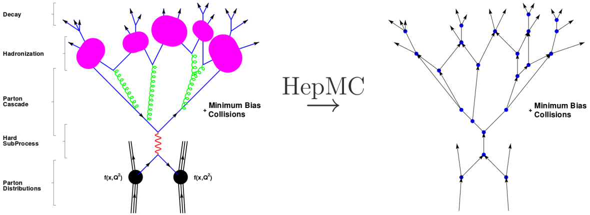
Analysis coverage, source:
https://rivet.hepforge.org/rivet-coverage

Rivet analysis

- Analyses as C++ classes
- To be written by users (analysers)
- Unique analysis identifier
  - Consist of experiment name, year of publication, and Inspire ID code
  - e.g. ALICE_2015_I1357424
- 3 core methods:
  - init(): histogram booking, variables initialization...
  - analyze(...): main event loop, events are analysed and histograms are filled here
  - finalize(): histogram scaling, divisions for the final data comparison

HepMC

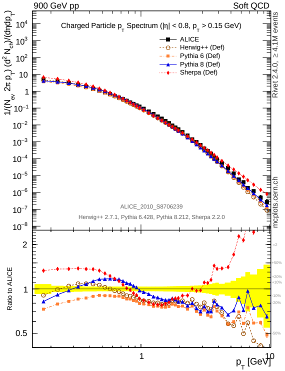
- HepMC:
  - Standarized event record library for High Energy Physics Monte Carlo generators and simulation, arXiv:1912.08005
  - A direct output of most modern generators
  - File format contains information about final state particles
  - Why? We need an interface between generator and analysis

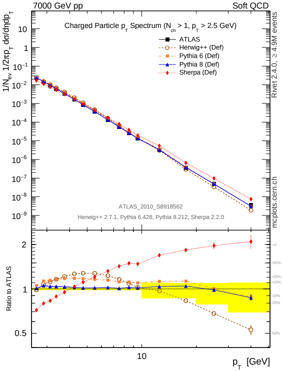


- Recent workshop:
  - HEPMC in Heavy Ion Collisions (7 June 2021)
  - https://indico.bnl.gov/event/10966/

MCPLOTS website

- MCPLOTS: a particle physics resource based on volunteer computing, arXiv:1306.3436
  - Enable anyone to quickly get an idea of how well a particular model/tune describes various data sets
  - Automatize Rivet creation of comparison plots
  - Online repository of plots: **http://mcplots.cern.ch**





How do we want to do that?

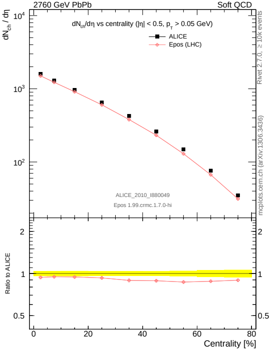
- How to make preservation possible?
  1. Develop tools that can be used for the analysis preservation
  2. Make these tools universal so that anyone within the HEP community can use them for their analysis
  3. Implement policies among collaborations allowing/enforcing them to follow the rules on what components to preserve and how to do that

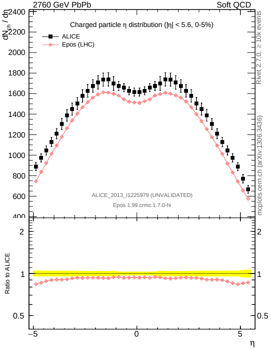
Rivet for heavy-ions

- Rivet originally developed for $pp/p\bar{p}$ and e^+e^- physics
- But in recent years this has changed
  - More interest from the heavy-ion community
  - Possibility to use Rivet as a tool for HI analyses
- Most common HI features already covered in Rivet 2.7+
  - see arXiv:1912.05451, arXiv:2001.10737
- Rivet released with:
  - Centrality calibration & selection using analysis options
  - Comparing and merging Rivet runs using reentrant finalize (e.g. for AA/pp ratio)
  - Flow observables, event mixing framework
  - A set of 20+ analyses using heavy-ion features

MCPLOTS for heavy-ions

- Originally provided results for pp and ee, but not for heavy-ions
- New **heavy-ion version** of MCPLOTS
  - Uses new Rivet version as a baseline
  - Enables to run analyses that use heavy-ion features described
- Integration ongoing
  - Preview version available here: **http://mcplots-alice.cern.ch**





Obstacles in heavy-ion analyses

- MC event production in pp:
  - low CPU time, low storage consumption
  - can easily run with FIFO logic
- MC event production in AA:
  - large CPU time, huge storage consumption (compared to pp)
  - difficult to run with FIFO logic

Event type	MC generator	CPU time / event	Storage / HepMC event
pp, 5.02 TeV, min. bias	Pythia 8 Monash Tune	$\approx$ 20 ms	$\approx$ 50 kB
Pb-Pb, 5.02 TeV, min. bias	Pythia 8 Monash Tune	$\approx$ 1 s	$\approx$ 5 MB

- Not all heavy-ion features available yet when using tools like Rivet
  - limited background subtraction
  - fits (e.g. for extracting femtoscopic radii)

How do we want to do that?

- How to make preservation possible?
  1. Develop tools that can be used for the analysis preservation
  2. Make these tools universal so that anyone within the HEP community can use them for their analysis
  3. Implement policies among collaborations allowing/enforcing them to follow the rules on what components to preserve and how to do that

ALICE policy

1. HepData submission policy
    - No HepData available at the 'ALICE preliminary' level
    - HepData tarball (yaml) prepared for the ArXiv submission
    - Remains ALICE internal until publication in the journal
    - HepData tarball pushed to website once published by the journal
2. Analysis rivetization
    - Current goal: after publication (once HEPdata is publicly available)
        - Many older analyses are still waiting to be rivetized
    - Encouraged: during paper preparation
        - User can benefit from the comparison to models

ALICE preservation effort

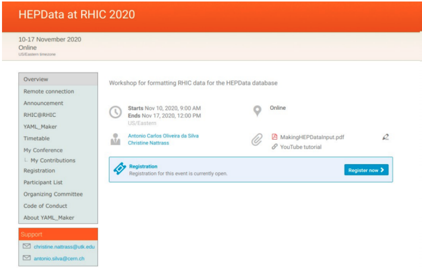
- The work was constrained to pp analyses due to limited support for heavy-ions until recently
- People in ALICE are working on rivetizing their analyses, including AA analyses
  - The trend is growing
  - A dedicated group with the objective of coordinating the analysis rivetization within the experiment
- ALICE developments
  - What we publish is reproduceable
  - Code is publicly available

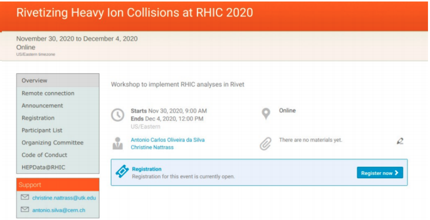
Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	e^+e^- (≥ 12 GeV)	e^+e^- (≤ 12 GeV)	Tevatron	RHIC	SPS	Other
Rivet provided:	12/197 = 6%	2/52 = 4%	0/79 = 0%	0/7 = 0%	0/1 = 0%	0/1 = 0%	0	0	0	3/303 = 1%	0	1/1 = 100%

HI analysis coverage, source:
https://rivet.hepforge.org/rivet-coverage-heavyiononly

Rivetization at RHIC

- Similar approach:
  - Use Rivet + HepMC
- Various challenges:
  - Missing features in Rivet
  - Data not in HepData
- Many activities:





Activities at RHIC

- How to get the work done, especially when people that measured something are long gone?
- Solution: Course-based Undergraduate Research experience
- Total number of involved students (up to now):
  - 28 students
  - 12 women
  - 9 minorities
  - 4 non-traditional

Poster session

"Rivet for heavy ions at RHIC: theory comparison and education"
Antonio Carlos Oliveira Da Silva (University of Tennessee - Knoxville) & Christine Nattrass (University of Tennessee (US))

Summary

- Preservation is essential
- The tools are there and their integration with the heavy-ion features is ongoing
  - First frameworks are there
  - More heavy-ion integration is expected
- Analysis preservation activities are visible within ALICE and RHIC
- Heavy-ion communities are working on implementing and standarizing their policies
- ALICE and RHIC are (and will be) pushing towards preservation of their analyses

References

- Rivet webpage: https://rivet.hepforge.org/
- Rivet user manual: arXiv:1003.0694
- Rivet version 3: arXiv:1912.05451
- Rivet for heavy-ions: arXiv:2001.10737
- MCPLOTS webpage: http://mcplots.cern.ch/
- MCPLOTS HI development webpage: http://mcplots-alice.cern.ch/

Thank you!

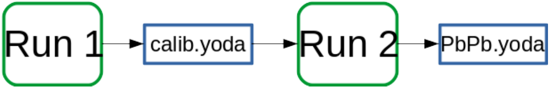
Backup

More about Rivet

- Rivet features
  - Provides a large set of experimental analyses (currently over 400...) useful for MC generator development, validation, and tuning
  - Analyses correspond to the actual paper results
  - Analysis algorithms in object-oriented C++
  - Analysis code separated $\rightarrow$ easy way to add a new one
  - Using the HepMC event format $\rightarrow$ independent of MC generators
  - Plotting based on YODA framework

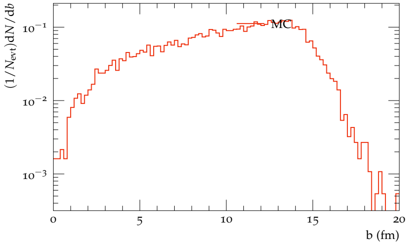
Centrality framework in Rivet

- Centrality framework: allows an analysis to cut on percentiles of single event quantities preloaded to that analysis
- Consist of:
  - A set of calibration analyses
  - Preloading calibration files as an input to the next run
  - Options for centrality to select the type of calibration to use
  - Centrality projection that allows to access centrality value
- Analysis options allow selection of centrality calibration method
  1. Calibration histogram from reference data
  2. Generated calibration histogram
  3. Impact parameter calibration histogram
  4. User-defined calibration histogram
  5. Generated centrality (available only with HepMC3)

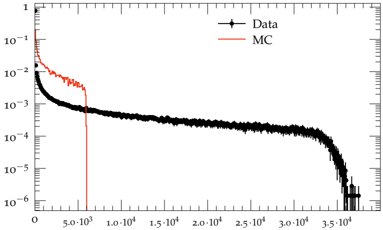


Centrality calibration

- Calibration analysis used to produce calibration files
- Example analysis: ALICE_2015_PBPBCentrality



Impact parameter distribution

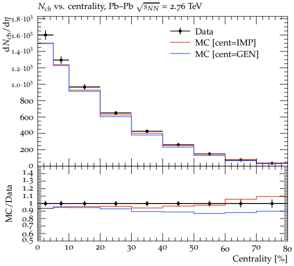


V0M multiplicity distribution

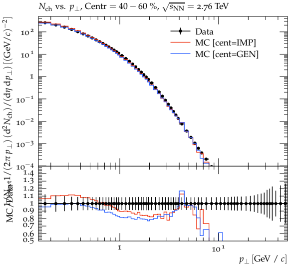
- New *preload* flag to preload histograms to the Rivet runs

Centrality selection with analysis options

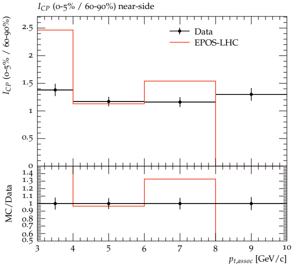
- Analysis options
  - Possible to select by user at runtime
  - Analysis can be run with different options
  - Centrality value accessible inside an analysis



Example from ALICE_2010_I880049



Example from ALICE_2012_I1127497



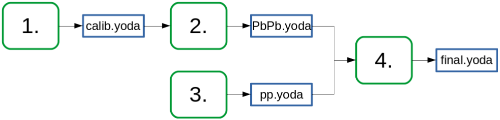
Example from ALICE_2012_I930312

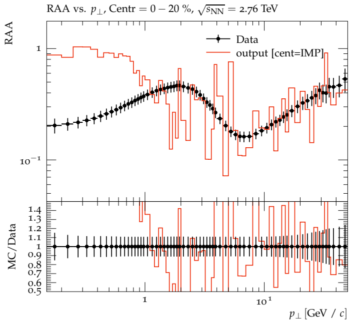
Reentrant finalize

- Postprocessing implemented in a form of 'reentrant finalize' method
  - Analyses produce 'RAW' histograms - saved in the output file before calling finalize() method
  - rivet-merge to call finalize() again, but this time with preloaded files from previous runs to get final plots
  - 'Reentrant' flag to mark analyses using reentrant finalize method

Nuclear modification factor

$$R_{AA} = \frac{\mathrm{d}N_{AA}/\mathrm{d}p_T}{<N_{coll}>\mathrm{d}N_{pp}/\mathrm{d}p_T}$$





ALICE_2012_I1127497 example

HI analyses list

With data:

1. ALICE_2010_I880049 (PbPb): Multiplicity
2. ALICE_2012_I930312 (PbPb): I_{AA}
3. ALICE_2012_I1127497 (PbPb): R_{AA}
4. ALICE_2012_I1126966 (PbPb): π, K, p spectra
5. ALICE_2013_I1225979 (PbPb): Charged multiplicity
6. ALICE_2014_I1243865 (PbPb): Multi-strange baryons
7. ALICE_2014_I1244523 (pPb): Multi-strange baryons
8. ALICE_2015_PBPBCentrality (PbPb): Calibration
9. ALICE_2016_I1394676 (PbPb): Charged multiplicity
10. ALICE_2016_I1419244 (PbPb): Multiparticle correlations (flow)
11. ALICE_2016_I1471838 (pp): Multi-strange baryons
12. ALICE_2016_I1507090 (PbPb): Charged multiplicity
13. ALICE_2016_I1507157 (pp): Angular correlations
14. ATLAS_2015_I1360290 (PbPb): Charged multiplicity
15. ATLAS_2015_I1386475 (pPb): Charged multiplicity + spectra
16. ATLAS_PBPB_CENTRALITY (PbPb): Calibration
17. ATLAS_pPb_Calib (pPb): Calibration
18. BRAHMS_2004_I647076 (AuAu): π, K, p spectra as function of

Without data:

1. ALICE_2015_PPCentrality: Calibration
2. BRAHMS_2004_CENTRALITY: Calibration
3. STAR_BES_CALIB: Calibration
4. MC_Cent_pPb_Calib: Calibration example
5. MC_Cent_pPb_Eta: Calibration example
6. MC_OPTIONS: Analysis options example
7. MC_REENTRANT: Reentrant finalize example

See list of all analyses here

HI analysis chain

1. Calibration run:

```
rivet -a ALICE_Calibration /path/to/PbPb.hepmc
```

2. Run with preloaded calibration file for PbPb beam:

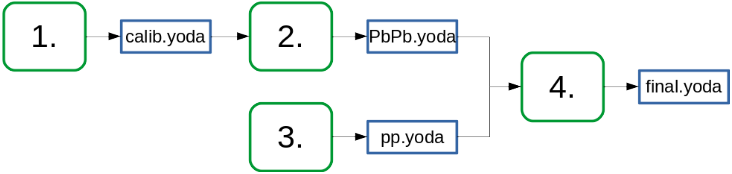
```
rivet -p /path/to/calib.yoda -a ALICE_2012_I123456:cent=IMP /path/to/PbPb.hep
```

3. Regular run for pp beam:

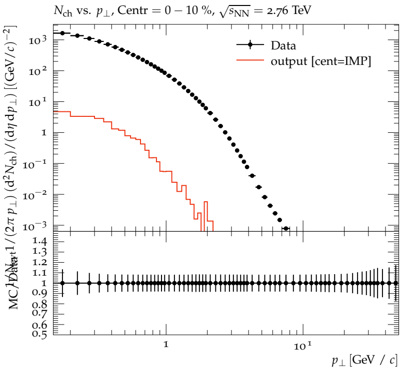
```
rivet -a ALICE_2012_I123456:cent=IMP /path/to/pp.hepmc
```

4. Postprocessing with reentrant finalize:

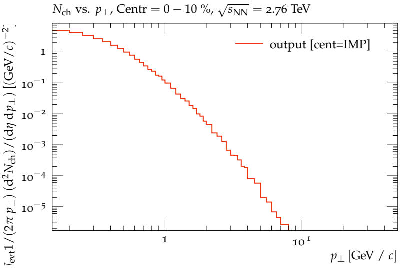
```
rivet-merge PbPb.yoda pp.yoda -o final.yoda
```



R_{AA} analysis



(a) PbPb results



(b) pp reference

- Impact parameter distribution used for calibration
- No HepMC data for pp reference
- JEWEL does not provide N_{coll} - that's why the difference between MC and experimental data

R_{AA} analysis

Nuclear modification factor

$$R_{\mathrm{AA}} = \frac{\mathrm{d}N_{\mathrm{AA}}/\mathrm{d}p_{\mathrm{T}}}{<N_{\mathrm{coll}}>\mathrm{d}N_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}}$$

$\langle N_{\mathrm{coll}} \rangle$ - average number of nucleon-nucleon collisions in a HI collision
$\mathrm{d}N_{\mathrm{AA}}/\mathrm{d}p_{\mathrm{T}}$ - transverse momentum distribution in HI collision
$\mathrm{d}N_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}$ - transverse momentum distribution in pp collision

- R_{AA} tells how heavy-ion events looks like in scale of pp collisions
- Both pp and heavy-ion collisions are required to calculate R_{AA}

YODA

- YODA (Yet more Objects for Data Analysis):
  - A histogramming toolkit developed as a lightweight common system for MC event generator validation analyses
  - The core histogramming system in Rivet

YODA logo