How to make HEPData input

A how-to guide designed for legacy analyses

There are many ways to do this. My goal here is to show you the quickest.

Video tutorial available here

Why upload to HEPData?

Centrally maintained database used by particle physics and LHC experiments

Reduces reliance on websites maintained by collaborations

Makes implementation of analysis in Rivet easier

Download and install YAMLMaker

Github repository: https://github.com/tkrobatsch/YAML_Maker

git clone https://github.com/tkrobatsch/YAML_Maker.git
cd YAML_Maker/
make

The YAML files for upload to HEPData are just formatted text files. YAML_Maker takes simple text files with the data and makes them into correctly formatted YAML files.

Shout out to undergraduate Tom Krobatsch!

Note that hepdata-converter is also available. It is dependent on ROOT6 and somewhat dependent on python versions. I have not found it as easy to use.

Format text files with data

PHENIX papers are here
Example:
PHENIX Phys. Rev. C 87, 034911 (2013)
arxiv:1208.2254
Inspire 1127262
PPG133

Data are available here

Header

p_T x-axis
4 Number of data sets
η/π^{0} y-axis - data set 1
η/π^{0} y-axis - data set 2
η/π^{0} y-axis - data set 3
η/π^{0} y-axis - data set 4
$0-20\%$ centrality label - data set 1
$20-60\%$ centrality label - data set 2
$60-93\%$ centrality label - data set 3
$0-93\%$ centrality label - data set 4
yes bins? **You want to use bins for Rivet implementation!**
none x statistical uncertainty (symmetric, asymmetric, none)
none x systematic uncertainty (symmetric, asymmetric, none)
3 Number of uncertainties?
p_{T} Uncorr Uncertainty 1 label
symmetric Unc. 1 type (symmetric, asymmetric, none)
p_{T} Corr Uncertainty 2 label
symmetric Unc. 2 type (symmetric, asymmetric, none)
p_{T} typeC Uncertainty 3 label
symmetric Unc. 3 type (symmetric, asymmetric, none)

Latex accepted, but this is often the cause of errors!

Format text files with data

PHENIX papers are here
Example:
PHENIX Phys. Rev. C 87, 034911 (2013)
arxiv:1208.2254
Inspire 1127262
PPG133

Data are available here

Body

```
x range y       yerr1       yerr2       yerr3
5-6     0.46194 0.026834    0.048098    0.0069292
6-7     0.41051 0.031577    0.042743    0.0061577
7-8     0.47024 0.039008    0.048961    0.0070536
8-9     0.55552 0.0472  0.052187    0.0083328
9-10    0.44789 0.053741    0.042076    0.0067184
10-12   0.51979 0.049027    0.04883 0.0077969
12-14   0.59078 0.073435    0.055498    0.0088617
14-16   0.28979 0.085438    0.027223    0.0043468
16-18   0.43909 0.13536 0.041248    0.0065863
***
5-6     0.51477 0.020629    0.053598    0.0077216
6-7     0.53073 0.023336    0.05526 0.0079609
7-8     0.50308 0.026865    0.052381    0.0075462
8-9     0.48272 0.032162    0.045347    0.0072407
9-10    0.5492  0.041658    0.051592    0.0082379
10-12   0.50214 0.040903    0.047172    0.0075322
12-14   0.47711 0.062042    0.04482 0.0071566
14-16   0.34917 0.10502 0.032802    0.0052376
16-18   0.54718 0.21836 0.051403    0.0082077
***
5-6     0.49549 0.025516    0.051591    0.0074324
```

Tab delimited

Indication of end of data set

Run YAML_Maker on all text files

./yaml_data 0 text/fig8.txt

The first argument is the debug level (0=quiet, >0=verbose)

The second argument is the text file.

Shout out to undergraduate Christal Martin, who formatted these text files!

submission.yaml

Top level file which tells HEPData what to read

```
comment: | Abstract: no line breaks and beware Latex-related errors!
  BNL-RHIC.  The PHENIX experiment has measured the production of neutral pions in Au+Au collisions at
$\sqrt{s_{NN}}$=200 GeV. The new data offer a fourfold increase in recorded luminosity, providing higher precision and a larger
reach in transverse momentum, $p_T$, to 20 GeV/c. The production ratio of $\eta/\pi^0$ is
0.46$\pm$0.01(stat)$\pm$0.05(syst), constant with p_T and collision centrality. The observed ratio is consistent with earlier
measurements, as well as with the p+p and d+Au values. The production of $\pi^0$ is suppressed by a factor of 5, as in earlier
findings. However, with the improved statistical precision a small but significant rise of the nuclear modification factor, $R_AA$,
vs $p_T$, with a slope of 0.0106$^(+0.0034)_(-0.0029)[GeV/c]^-1$, is discernible in central collisions. A phenomenological
extraction of the average fractional parton energy loss shows a decrease with increasing p_T. To study the path length
dependence of suppression, the $\pi^0$ yield was measured at different angles with respect to the event plane; a strong
azimuthal dependence of the $\pi^0 R_AA$ is observed. The data are compared to theoretical models of parton energy loss as
a function of the path length, L, in the medium. Models based on pQCD are insufficient to describe the data, while a hybrid
model utilizing pQCD for the hard interactions and AdS/CFT for the soft interactions is consistent with the data.
additional_resources:
  - {location: "https://www.phenix.bnl.gov/phenix/WWW/info/data/ppg133_data.html", description: "web page with data points"}
--- Figure delimiter
name: "Figure 5"
description: 'Invariant yields of neutral pions, all centralities'
keywords:
  - {name: reactions, values: [AU AU --> NEUTRAL X]}
  - {name: cmenergies, values: [200.0]}
data_file: fig5.yaml
---
```

HEPData.net

Need to create an account to test

Click on link to sandbox. Package all YAML files in a tar ball or zip file.

I recommend iteratively adding each table. Errors are not always useful. The most common errors are related to Latex. Picky about white space.

For legacy analyses, I advocate minimizing time debugging formatting and focusing on getting something accurate and clear up.

Upload an archive to the sandbox

This is a private upload area.

Upload an archive (**.zip**, **.tar**, **.tar.gz**, **.tgz**) containing YAML files formatted according to these guidelines. An example submission archive is available here. You can validate your YAML files offline using this script.

We also accept a **single YAML** file containing all of the submission data. For records uploaded to the old HepData site, this format can be obtained automatically by appending "/yaml" to the old record URL.

Alternatively, upload a single text file with extension **.oldhepdata** containing the "input" format that was used for data submissions from the old HepData site (see sample).

1208.2254.tgz
Choose file

Upload and Process

Your sandbox submissions

BNL-RHIC. We report high statistics measurements of inclusive charged hadro...
2020-04-07 12:32PM

BNL-RHIC. We report high statistics measurements of inclusive charged hadro...
2020-04-02 19:32PM

BNL-RHIC. We report high statistics measurements of inclusive charged hadro...
2020-04-02 19:30PM

BNL-RHIC. We report high statistics measurements of inclusive charged hadro...
2020-04-02 19:25PM

Validation error encountered

A number of errors were encountered by our validation code. Our validation ensures that your submission files match the format described here. You can validate your YAML files offline using this script.

Please fix your submission and reupload via the associated record page.

fig11.yaml

error — Uncertainties should not all be zero in {'errors': [{'symerror': 0, 'label': 'p_{T} Uncorr'}, {'symerror': 0, 'label': 'p_{T} Corr'}, {'symerror': 0, 'label': 'Global'}], 'value': 0}

fig21.yaml

error — There was a problem parsing the file. while parsing a block mapping in "/opt/hepdata/var/data/1588150448/1588149686/1208.2254/fig21.yaml", line 2, column 3 did not find expected key in "/opt/hepdata/var/data/1588150448/1588149686/1208.2254/fig21.yaml", line 5, column 5

Additional Resources

Abstract (data abstract)

BNL-RHIC. The PHENIX experiment has measured the production of neutral pions in Au+Au collisions at $\sqrt{s_{NN}}$=200 GeV. The new data offer a fourfold increase in recorded luminosity, providing higher precision and a larger reach in transverse momentum, p_T, to 20 GeV/c. The production ratio of η/π^0 is 0.46±0.01(stat)± 0.05(syst), constant with p_T and collision centrality. The observed ratio is consistent with earlier measurements, as well as with the p+p and d+Au values. The production of π^0 is suppressed by a factor of 5, as in earlier findings. However, with the improved statistical precision a small but significant rise of the nuclear modification factor, R_AA, vs p_T, with a slope of $0.0106^{(}+0.0034)_{(}-0.0029)[GeV/c]^{-}1$, is discernible in central collisions. A phenomenological extraction of the average fractional parton energy loss shows a decrease with increasing p_T. To study the path length dependence of suppression, the π^0 yield was measured at different angles with respect to the event plane; a strong azimuthal dependence of the $\pi^0 R_AA$ is observed. The data are compared to theoretical models of parton energy loss as a function of the path length, L, in the medium. Models based on pQCD are insufficient to describe the data, while a hybrid model utilizing pQCD for the hard interactions and AdS/CFT for the soft interactions is consistent with the data.

Download All

Filter 10 data tables

Figure 8
$\Eta/ \pi^0 ratios

Figure 11
π^0 nuclear modification factors

Figure 20-1
pi^0 nuclear modification factors vs reaction plane, $\Delta\phi = 0^\circ - 15^\circ$

Figure 20-2
pi^0 nuclear modification factors vs reaction plane, $\Delta\phi = 15^\circ - 30^\circ$

Figure 20-3
pi^0 nuclear modification factors vs reaction plane, $\Delta\phi = 30^\circ - 45^\circ$

Figure 20-4
pi^0 nuclear modification factors vs reaction plane, $\Delta\phi = 45^\circ - 60^\circ$

Figure 20-5
pi^0 nuclear modification factors vs reaction plane, $\Delta\phi = 60^\circ - 75^\circ$

Figure 20-6
pi^0 nuclear modification factors vs reaction plane, $\Delta\phi = 75^\circ - 90^\circ$

Figure 21

Centrality dependence of p_T ingetrated $R_{AA}(\Delta\phi)$

cmenergies

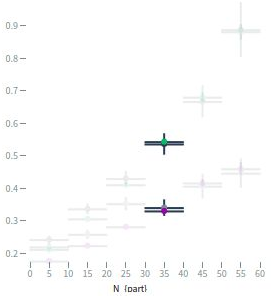
200.0

reactions

AU AU --> NEUTRAL X

N_{part}	$(6 < p_T < 10GeV/c), (0 < \Delta\phi < 15)$	$(6 < p_T < 10GeV/c), (75 < \Delta\phi < 90)$	$(pT > 1$
N_{part}	$R_{AA}(p_T, \Delta\phi)$		
0 - 10	0.2123 ±0.002549 Stat ±0.0017 Syst	0.1766 ±0.0022 Stat ±0.0018 Syst	0.2428 ±(
10 - 20	0.3063 ±0.003552 Stat ±0.002 Syst	0.2241 ±0.0029 Stat ±0.0021 Syst	0.3371 ±(
20 - 30	0.4113 ±0.005015 Stat ±0.0031 Syst	0.2821 ±0.0039 Stat ±0.0031 Syst	0.4306 ±(
30 - 40	0.5439 ±0.007235 Stat ±0.0047 Syst	0.3308 ±0.0053 Stat ±0.0046 Syst	0.5368 ±(
40 - 50	0.6809 ±0.0109 Stat ±0.0092 Syst	0.4168 ±0.008 Stat ±0.0092 Syst	0.6676 ±(
50 - 60	0.8819 ±0.01744 Stat ±0.016 Syst	0.4605 ±0.011 Stat ±0.017 Syst	0.8886 ±(

Visualize



Sum errors | Log Scale (Y)

Deselect variables or hide different error bars by clicking on them.

Variables

$R_{AA}(p_T, \Delta\phi)$
N_{part};$(6 < p_T < 10GeV/c), (0 < \Delta\phi < 15)$
Summed error

Tips

This is not the time or place to express your creativity!

Start with a minimal working example. Add data iteratively and test often using the sandbox.

Use bins - it will pay off during Rivet implementation

Don’t worry too much about Latex formatting

Sample files: https://drive.google.com/drive/folders/1ZFqggDSwQmH2XjxXKqk5ro327wKYgCqd?usp=sharing

Note: the final step is for the collaboration’s designated HEPData submitter to approve the file.