

# The PHENIX DAP update

- Workshops
- HEPData/GitHub
- Website updates
- Zenodo, Keywords and Conferences
- REANA plans
- Status of PHENIX access to the Open Data portal

Maxim Potekhin  
(BNL, NPPS)  
12/10/2020

# Workshops

- Great many thanks from all of us to Christine for organizing these events for the benefit of the RHIC community
- Not lectures but hands-on workshops/tutorials on HEPData and Rivet
- HEPData: <https://indico.bnl.gov/event/8843/timetable/> (11/10 & 11/17 2020)
  - Org Committee: Christine Nattrass, Antonio Da Silva, Lauren Kasper, Andi Mankolli
- Rivet: <https://indico.bnl.gov/event/8840/timetable/> (11/20-12/4 2020)
  - Org Committee: Christine Nattrass, Antonio Da Silva, Christian Bierlich, Leif Lönnblad

# Rivet, HEPData, YODA

- Integration with InspireHEP (ID)
- These topics are related since data exported from HEPData in the format called “YODA” are a part of input for Rivet. YODA is a set of analysis and histogramming classes, and the core of histogramming in Rivet
- There are *some* compatibility issues when exporting the data:
  - [https://indico.cern.ch/event/808138/contributions/3364224/attachments/1851728/3040138/watt\\_hepdata\\_may2019.pdf](https://indico.cern.ch/event/808138/contributions/3364224/attachments/1851728/3040138/watt_hepdata_may2019.pdf)
  - [https://conference.ippp.dur.ac.uk/event/875/contributions/4767/attachments/3898/4443/HepData\\_Rivet\\_2020-01.pdf](https://conference.ippp.dur.ac.uk/event/875/contributions/4767/attachments/3898/4443/HepData_Rivet_2020-01.pdf)
- Discussions with Christine re: creating functional pages for our site (next slide)
- Interesting to see the scope of what’s possible with Rivet in PHENIX
- Caveat: Rivet operates on fairly final data products which insulates it from the nitty-gritty of reconstruction and many analysis steps

# Rivet content on the site

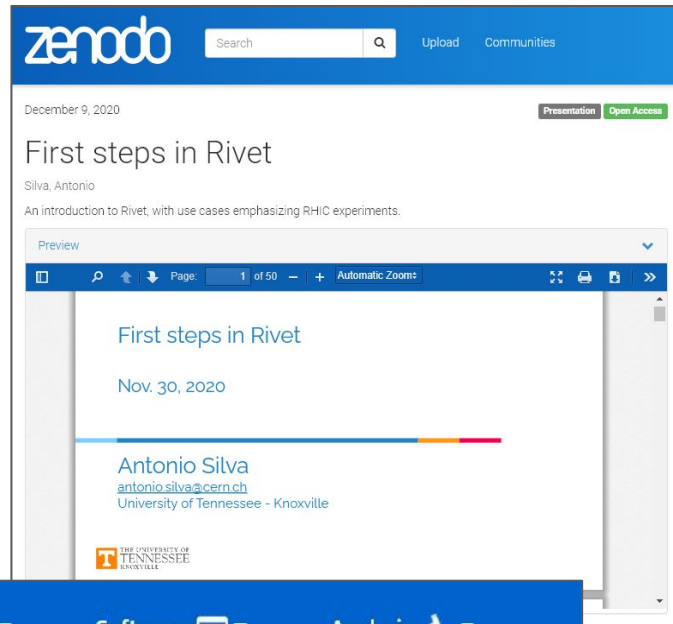
- Updated the Rivet page (in “Analysis”)
- Uploaded slides to Zenodo
- Added “rivet” to the keyword list
- Work in progress, lots to do






## Simulation (3 items)

Keyword	Description
<a href="#">hijing</a>	HIJING - a type of Event Generator
<a href="#">pisa</a>	PISA - the PHENIX Integrated S
<a href="#">pythia</a>	PYTHIA - a type of Event Generator

## Misc Software (2 items)

Keyword	Description
<a href="#">rivet</a>	Robust Independent Validation
<a href="#">software</a>	Misc software



[Experiment](#)  [Results](#)  [Detectors](#)  [Software](#)  [Analysis](#) 

---

## Rivet

The Rivet toolkit (Robust Independent Validation of Experiment and Theory) is currently under evaluation by PHENIX.

### Links

[Rivet website](#)

### Tutorials

- DOI [10.5281/zenodo.4313877](https://doi.org/10.5281/zenodo.4313877) First steps in Rivet (A. Silva)

# New Repository for HEPData

- HEPData work is progressing, active participants clone/fork the existing repo
  - Six active people in the past week
- The default “documentation” repo is an umbrella for all sorts of materials
- It has grown to a large size - thanks in part to Stacyann’s valuable effort to get our theses’ materials organized in conjunction with Zenodo uploads
- But the repo doesn’t scale and working with HEPData has become more cumbersome than it should be
- Solution: create a dedicated repo owned by PHENIX for clean and efficient access
  - It has been done: <https://github.com/PhenixCollaboration/hepdata>
  - Uniform naming scheme for the folders i.e. ppgXXX
- Links updated on the website

# Repo contents: 25 submission packages+examples

PhenixCollaboration / hepdata

Unwatch 2

<> Code ⓘ Issues 🔗 Pull requests ⚙️ Actions 📁 Projects 📖 Wiki 🔒 Security 📊 Insights ⚙️ Settings 📦 3.52 MB

master 1 branch 0 tags

Go to file Add file Code

**buddhasystem** Added to README, tweaked ppg119 8a76f08 35 minutes ago 9 commits

examples	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg055	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg071	Adding ppg071	4 days ago
ppg079	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg083	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg090	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg092	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg095	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg113	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg116	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg119	Added to README, tweaked ppg119	35 minutes ago
ppg129	Migrating the content from "documentation/assets" to this new r...	4 days ago
ppg132	Migrating the content from "documentation/assets" to this new r...	4 days ago

**About**

A repository for managing the PHENIX materials submitted to the HEPData portal

Readme

**Releases**

No releases published  
[Create a new release](#)

**Packages**


No packages published  
[Publish your first package](#)

**Contributors** 2

**buddhasystem** Maxim Potekhin

**cnattras**

# Our HEPData content (46 items right now)



Submit Sandbox About Submission Help Dashboard Logout


phenix

PHENIX Search Reset search Advanced

Max results Sort by Reverse order

Showing 25 of 46 results

Date



Collaboration

PHENIX

46

Subject\_areas

nuc-ex

33

hep-ex

17

Phrases

Inclusive

28

Proton-Proton Scattering

17

Electron production

16

Rapidity Dependence

13

Transverse Momentum Dependence

10

Next 3 Show All

Reactions

AU AU -> CHARGED X

7

Centrality dependence of charged particle multiplicity in Au - Au collisions at  $\sqrt{s(NN)} = 130$  GeV

The PHENIX collaboration Adcox, K. ; Adler, S.S. ; Ajitanand, N.N. ; *et al.*

Phys.Rev.Lett. 86 (2001) 3500-3505, 2001.

Inspire Record 539140 DOI 10.17182/hepdata.50270

We present results for the charged-particle multiplicity distribution at mid-rapidity in Au - Au collisions at  $\sqrt{s(NN)} = 130$  GeV measured with the PHENIX detector at RHIC. For the 5% most central collisions we find  $dN_{ch}/d\eta_{p+0} = 622 \pm 1(stat) \pm 41(syst)$ . The results, analyzed as a function of centrality, show a steady rise of the particle density per participating nucleon with centrality.

1 data table match query

Table 1 130 GeV  $\sqrt{s}$  per nucleon-nucleon collision.  $N(C=N_{NUCLEONS})$  and  $N(C=N_{COLLISIONS})$  are the number of participating nucleons and binary collisions. The statistical errors are negligible and only systematic errors are quoted.  $COL(NAME=CENTRALITY)$  is centrality.

Dilepton mass spectra in p+p collisions at  $\sqrt{s} = 200$  GeV and the contribution from open charm

The PHENIX collaboration Adare, A. ; Afanasiev, S. ; Aidala, C. ; *et al.*

Phys.Lett.B 670 (2009) 313-320, 2009.

Inspire Record 778611 DOI 10.17182/hepdata.73669

The PHENIX experiment has measured the electron-positron pair mass spectrum from 0 to 8 GeV/c<sup>2</sup> in p+p collisions at  $\sqrt{s}=200$  GeV. The contributions from light meson decays to  $e^+e^-$  pairs have been determined based on measurements of hadron production cross sections by PHENIX. They account for nearly all  $e^+e^-$  pairs in the mass region below 1 GeV/c<sup>2</sup>. The  $e^+e^-$  pair yield remaining after subtracting these contributions is dominated by semileptonic decays of charmed hadrons correlated through flavor conservation. Using the spectral shape predicted by PYTHIA, we estimate the charm production cross section to be  $544 \pm 59(stat) \pm 142(syst) \pm 200(model) \mu b$ , which is consistent with QCD ...

2 data tables match query

Table 1 Differential charm cross section at mid rapidity An additional  $\sim 39.5$  microbarn error, due to the validity of the model used to extrapolate the data, is not included The contribution from beauty estimated to be 3.7 microbarn, has been subtracted. The c->e branching ratio used was  $9.5 \pm 1.0\%$ .

Table 2 Total charm cross section An additional systematic error of  $\sim 200$  microbarn, due to the validity of the model used to extrapolate the data, is not included. To obtain the total charm cross section, the differential charm cross section has been extrapolated to the whole rapidity range, using a HVQMNR rapidity distribution with aCTEQ8M PDF.

# HEPData experience

- HEPData: the devil is in the detail. Progress has been made with users creating submission packages, but errors are still common and **take time to fix**
  - Quality of typeset (LaTeX) in various places on the page e.g. abstract, figure captions
  - More importantly, consistency of the numerical data components
  - Consistency (and availability) of thumbnails and main images needs to be checked
- Many careful “sandbox” checks are necessary to QA the materials
- In one case we had to update 3 times to get everything right (we want to avoid this, of course)
- Discovered that an older item (ppg071) needs some improvement, work in progress - thanks Krista!



# HEPData: website updates etc

- Finally we started keeping track of the Inspire IDs as a part of a submission, as well as the designated IRC member
- Caveat: *for smooth operation the reviewer needs to be registered on the portal under the same email as was used when the submission is defined*
- Performed more updates of the HEPData policy page based on recent work and communication with participants of the effort
- Keeping track of progress in the README file in the repository


## List of completed uploads


055, 079, 090, 092, 095, 113, 116, 129, 132, 133, 138, 146, 174, 182, 227, 228


## In progress/recently finished


- 182: A.Khatriwada (prep, reviewer): [khatriwada.ajeeta@gmail.com](mailto:khatriwada.ajeeta@gmail.com) - note: finalized
- 119: K.Smith (prep), Matt - [matthew.g.wysocki@gmail.com](mailto:matthew.g.wysocki@gmail.com) (reviewer) - note: pending the reviewer's decision
- 071: K.Smith (prep), Tony - [afrawley@fsu.edu](mailto:afrawley@fsu.edu)(reviewer) - note: existing HEPData item being reviewed for completeness


# HEPData: policy page updated (again)

Experiment 

Results 

Detectors 

Software 

Analysis 

By the policy established by the PHENIX IB, every paper containing tables and/or plots must be accompanied by a HEPData-compliant data package containing the tables and/or plots data before it is approved for publication. Please see the official policy document (sec. IV.iv):

- [PHENIX Policies for Publication and for Other Dissemination of Results](#)

Specifically:

"The IRC shall be empowered to adjudicate disagreements on the details of the paper. A near-consensus shall be a pre-requisite for the submission of the paper to the preprint arXiv and journal, but only after the IRC has certified that data tables are available in proper format for later submission to HEP data."

### The Format

The data package prepared for submission to HEPData must conform to the specific format required by the HEPData portal - please see the [HEPData documentation](#) for details of the requirements. There is a useful [collection of tips](#) on the HEPData site, please peruse it. In addition, the DAP team created [a few simple examples](#) kept in the PHENIX repository on GitHub to illustrate basic features and options of the HEPData format. Beginners are encouraged to experiment with these examples by using the "sandbox" feature of the HEPData Portal (see [Appendix B](#) below).

The basic idea of how a submission is structured is as follows. Data contents of each item included in the package (e.g. a plot) are described in a corresponding individual file formatted as YAML (e.g. if there are 5 plots in the paper you are expected to provide 5 YAML data files). In addition, a special YAML file *submission.yaml* describes the submission as a whole e.g. provides the names of the data and optional image files, **list of keywords** etc. It also contains an abstract (typically imported as LaTeX from the publication material); unfortunately, not every LaTeX feature will work correctly on HEPData and the output will need to be checked (see the "sandbox" reference below).

Since YAML allows comments - lines starting with a "#" sign - it is very easy to add any sort of extra information to *submission.yaml* that may be helpful for communication with members of the Collaboration, reviewers and for the workflow of the submission process in general. For example, it is necessary to provide the **Inspire ID** of the paper for the HEPData submission. It can be placed in a comment line. Also, including the PHENIX-internal **PPG identifier** is highly recommended as it reduces the chances of human error and facilitates communication. Both Inspire ID and the PPG identifier can be easily incorporated in the comment lines of the *submission.yaml* file mentioned above (i.e. in lines of text starting with "#"). There can be any number of comment lines. Including information about the designated reviewer (member of the IRC for the paper) as an additional comment line is encouraged. The following pattern of the top of the *submission.yaml* file may help illustrate this:

```
# PPG999
# InspireHEP: 99999
# Reviewing IRC member: M.Phenix mphenix@bnl.gov
```

This is not to be confused with the *comment* attribute of the YAML file which almost always contains the abstract of the published paper, typically typeset in LaTeX:

# HEPData content

- HEPData: naively, items simply reflect contents of papers
- What about supporting materials/data?
- There is no hard and fast policy suggested or enforced on the site
- In fact, different preparers can have different approaches
- Discussions took place lately
- Do we need to create basic guidelines and put them on the web page?

# Keywords and Conferences

- Thanks to Gabor for reviewing Zenodo uploads and normalizing the list of keywords, the web page has been updated correspondingly
- Added Zimányi School 2020 to the conference page, seeded the Zenodo uploads for this conference with a PHENIX overview talk (keyword: zs20)

**BROOKHAVEN**  
NATIONAL LABORATORY

(my) PHENIX latest result overview  
light hadrons in p+A and A+B

**ZIMÁNYI SCHOOL 2020**

28th ZIMÁNYI SCHOOL  
WINTER WORKSHOP  
ON HEAVY ION PHYSICS  
December 7-11, 2020  
Budapest, Hungary

Takao Sakaguchi  
Brookhaven National Laboratory

2020/12/8

Sakaguchi © Zimanyi 2020

1

(11.5 MB)

Size

See more details...

Indexed in

**OpenAIRE**

Publication date:  
December 9, 2020

DOI:  
DOI: [10.5281/zenodo.4313181](https://doi.org/10.5281/zenodo.4313181)

Keyword(s):

rhic phenix small systems jet light hadron raa  
au+au cu+au d+au u+u J/ψ kaon neutron  
proton flow spin asymmetry zs20

Meeting:  
[Zimányi School 2020](#)

# REANA revisited

- A couple of slides presented in the DAP meeting on 10/01/2020
  - See links to misc presentations in the slides from Oct.1 for more info
- <http://reanahub.io/> “Reproducible Research Data Analysis Platform”
- Structured, unambiguous description of analysis workflows
  - There are two available workflow languages/schemas to choose from
  - Used in real-life complex analyses
  - **Software environment (not just code) packaged as an image**
  - Needs to run as a container on a cluster, currently at CERN, an option for local installation at BNL TBD
  - Helpful GUI

# REANA - “Hello World” example: the description


```
version: 0.3.0
inputs:
  files:
    - code/helloworld.py
    - data/names.txt
  directories:
    - workflow/cwl
  parameters:
    input: workflow/cwl/helloworld-job.yml
workflow:
  type: cwl
  file: workflow/cwl/helloworld.cwl
outputs:
  files:
    - outputs/greetings.txt
```



```
sleeptime: 0
helloworld:
  class: File
  location: code/helloworld.py
inputfile:
  class: File
  location: data/names.txt
```

# REANA - an example of a real-world analysis

```
version: 0.4.0
inputs:
  files:
    - config/geantSim_TrackerPerformance.py
    - config/single_particle_trackFits.py
    - script/numHitsPerTrack.C
    - script/plot_single_particle_resolutions.py
  parameters:
    events: 5000
    seed: 0123456
    particle: 13
    etamin: 0
    etamax: 6
    pt: 1000 2000 5000 10000 100000 1000000 10000000
```



```
workflow:
  type: serial
  specification:
    steps:
      - environment: 'gitlab-registry.cern.ch/vavolkl/fcc-ubuntu:latest'
      commands:
        - fccrun.py config/geantSim_TrackerPerformance.py
          -N $events -s $seed --outName muons_for_seeding_discrete_pt.root
          --singlePart --particle $particle --etaMin $etamin --etaMax $etamax
          --discretePt --pt $pt
          --pathToDetector /usr/local/
          | tee simulation.log 2> simulation.err
        - root -b 'script/numHitsPerTrack.C("muons_for_seeding_discrete_pt.root")'
          | tee plot-tracker-hits.log 2> plot-tracker-hits.err
        - fccrun.py config/single_particle_trackFits.py
          --inputfile muons_for_seeding_discrete_pt.root
          --outputfile single_particle_resolutions.root
          | tee fit.log 2> fit.err
        - python script/plot_single_particle_resolutions.py
          single_particle_resolutions.root
          | tee plot-reconstructed-tracks.log 2> plot-reconstructed-tracks.err
```



Image

# REANA - another complex example

```
version: 0.3.0
inputs:
  files:
    - data.txt
    - AliAnalysisTaskEx01.cxx
    - AliAnalysisTaskEx01.h
    - runEx01.C
    - plot.C
    - OCDB.root
    - fix-env.sh
  parameters:
    data_location: http://opendata.cern.ch/record/1102/files/assets/alice/2010/LHC10h/000139038/ESD/0003/AllESDs.root
  workflow:
    type: serial
    specification:
      steps:
        - environment: 'reanahub/reana-env-aliphysics:vAN-20180614-1'
          commands:
            - mkdir data/ && wget -O data/AllESDs.root ${data_location}
            - source ./fix-env.sh && root -b -q './runEx01.C' | tee run.log
            - source ./fix-env.sh && root -b -q './plot.C' | tee plot.log
      outputs:
        files:
          - AnalysisResults.root
          - plot_pt.pdf
          - plot_eta.pdf
```





# REANA: plans

- REANA is a complex system, need to evaluate cost-benefit for PHENIX
  - Learning curve is a little steep
  - State of the art, well supported
  - Part of a wider question of the DAP priorities
  - Potential for cross-experiment aspect ???(sPHENIX, STAR, EIC)??? - recent Rivet/HEPData workshops are encouraging
- Ran a few simple tests on the CERN cluster, need to develop expertise
- Since PHENIX analysis are BNL-local due to various dependencies, the task is to build images that can run locally (AFS etc)
- Currently looking at Singularity as the main BNL framework to run containers
- If enough progress is made within weeks (perhaps months due to complexity) a request may be made to SDCC to create a local REANA cluster

# Open Data (the CERN portal)

- Since the last DAP meeting we have received storage allocation:  
[/eos/opendata/phenix/upload/](#)
- I have yet to test uploads, on my to-do list, will report

# Summary

- Steady progress with HEPData
  - Dedicated repository created
  - Website pages updated with latest instructions
  - A few packages are in the pipeline
- Two recent helpful workshops on HEPData and Rivet (C.Nattrass)
- Keyword updates, additions
- Good progress with conference material uploads to Zenodo (G.David)
- Thesis uploads and archiving on GitHub (Stacyann)
- REANA - initial exploration done, evaluation for PHENIX to follow
- The CERN Open Data portal access: storage defined, testing to be done