HowTo manage HEP data files at the new HEPData website

Tutorial for authors as Reviewers of ALICE HEP data submissions

20-03-2017

BASICS

Basics to submit to ALICE HEP Data to the new HEP data website: hepdata.net

- **ALICE Coordinator** for HEPData submissions: ALICE Webmaster
- **Uploader** of each submission: ALICE Webmaster
- **Reviewer of each submission**: PC Chair (designed as Reviewer by the webmaster)

Important URLs:

https://www.hepdata.net .... ← New HEPData website
https://www.hepdata.net/login/ .... ← Login page
https://www.hepdata.net/dashboard .... ← Own submissions
https://www.hepdata.net/submission .... ← Submission HELP
https://www.hepdata.net/record/sandbox .... ← Testing (tarballs or individual tables)

SUMMARIZING STEPS

1. **Creating of the submission record**
   - Create new submission
   - Assign “Reviewer”
   - Upload tarball previously tested by the PC Chair
   - Notify “Reviewer”

WEBMASTER AS COORDINATOR

2. **Review and approve the HEP data file**
   - Mark each table as having “passed”

PC CHAIR AS REVIEWER

3. **Finalise the submission**
   - Finalize submission from its dashboard

WEBMASTER AS COORDINATOR

4. **Replace the HEP Data tarball by the submitted YAML version in the paper node**

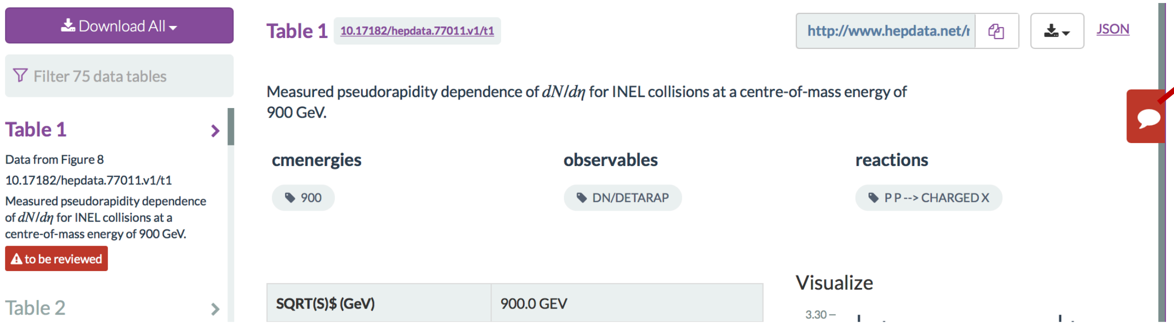
PC CHAIR AS REVIEWER

STEPS

PC CHAIR AS REVIEWER

1) The designed **Reviewer** should receive an e-mail with a link which assigns the appropriate privileges. Usually an email is also sent by the webmaster with detailed instructions about the procedure.

2) The **Reviewer** should mark each table as having “**Passed**” before proceeding with the next steps.

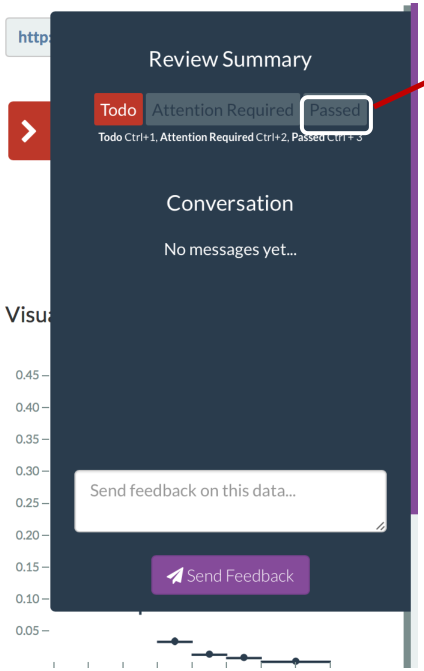


Data from Figure 8
10.17182/hepdata.77011.v1/t1
Measured pseudorapidity dependence of $dN/d\eta$ for INEL collisions at a centre-of-mass energy of 900 GeV.

SQRT(S)$ (GeV)	900.0 GEV

Review Summary:
Located to the right of each data table to approve, reject of have a discussion about the data table and its associate plot.

“**Passed**”



No messages yet...

Send feedback on this data...

3) The **Reviewer** sends an e-mail to the webmaster when all tables are marked as having “**Passed**”.

STEPS

FINALISING HEP DATA SUBMISSION

1) When tables are marked as “**Passed**”, the **Reviewer** should notify it by e-mail to the webmaster as coordinator.

2) The **Coordinator** should finalise the submission from its dashboard.

3) When the entire submission is published and made searchable in HEPData, the Coordinator sends an email notifying PC-Chair.

4) The HEP Data file should be replaced by the tarball final YAML version in the paper node at the ALICE internal pages.

YAML submission archive

HEP data submission involve the upload of archives (.zip, .tar, tar.gz) that specify the data associated with a publication

Example files and submission format

- Generic example of YAML submission can be found here: https://github.com/HEPData/hepdata-submission

- Examples of ALICE HEP data files based on YAML format can be seen by downloading records from www.hepdata.net and the tar-balls from their internal node in *aliceinfo.cern.ch/ArtSubmission*

Examples of ALICE HEP Data recently submitted to the new hepdata.net site

HEP Data record	Tarball from ALICE internal node
hepdata.net/record/76790	/files/draft/jcastill/2017-Apr-23-hepdata-final-id2688.tar.gz
hepdata.net/record/77011	/files/draft/aalkin/2017-Mar-28-HEData-yaml.tar_.gz
hepdata.net/record/77057	/files/draft/auras/2017-Apr-23-subm-final202.tar.gz
hepdata.net/record/77274	/files/draft/arossi/2017-Apr-13-hepdata-final-id1149.tar.gz
hepdata.net/record/77359	/files/draft/dstocco/2017-Apr-12-hepdata-final-id1815.tar.gz

YAML submission archive

The tar-ball file structure

- submission.yaml

Main file for a submission.

The whole submission is described by detailing the data files to be uploaded, their name and description, and their associated analysis files and code.

Required information: *name*, *description*, *keywords*, and the reference to the *data_file*.

Additional information: *location* of data within the paper, *license*, *additional resources* (e.g. code or ROOT files used to create these data)

- YAML data files

Data files encoded as YAML are defined in two parts which describe:

a) The independent variables (the x-axis in a plot).

b) The dependent variables (the things you are measuring, e.g. the y-axis in a plot)

In each you can define:

- *header*: the column name
- *values*: the rows in your table

For the independent variables you can define:

- *qualifiers*: extra metadata describing the measurement, i.e. energy, reaction type, possible kinematic cuts

- Additional files

Other files specified in the **submission.yaml** in fields like *location*, *license*, *additional_resources*, etc.

YAML submission archive

Uncertainties

There are two main classes of uncertainties that can be encoded:

- Symmetric errors: allow to specify plus and minus errors using one value, e.g.

```
symerror: 0.4
```

- Asymmetric errors: allow both, plus and minus errors to be explicitly encoded, e.g.

```
asymerror: {plus: 0.4, minus: -0.3}
```

Note that plus and minus can refer to «up» and «down» variations of the source of uncertainty.

YAML submission archive

Common mistakes and syntax errors

- **Scape special characters**

Some characters in YAML need to be escaped. The most problematic are ‘:’ and ‘-’.
So, the whole string where they are used needs to be quoted.

- **Ensure spaces after colons**

`{sumerrors:0.4, label:stats}` → gives you an error
`{sumerrors: 0.4, label: stats}` → works nicely!

- **Indentation and punctuation is very important in YAML!!!!**

- **The submission.yaml should not finish with an uncommented ‘---’ line, which defines an empty YAML document and currently causes an error.**

YAML submission archive

Testing the tar-ball before the official submission

For testing purposes, any user may use the Sandbox link at the new HEP data site: https://www.hepdata.net/ using the CERN credentials at “Log in with CERN” button. The steps are:

- Sign in https://www.hepdata.net/ using the CERN credentials at “Log in with CERN” button.
- Click on the Sandbox link to the top-right corner menu.
- Click “Choose file” and upload the tar-ball (tar.gz) containing the YAML files.

Since Sandbox and the submission process are unrelated it can be used to test data sets without any risk. In Sandbox you can test either, the YAML *data_files* individually or the whole submission by uploading the tarball (tar.gz).

YAML submission archive

Tips to start writing YAML data files

Sometimes the **.oldhepdata* file work well when is submitted to the new hepdata.net, but sometimes it doesn't.
In both cases you should prepare the tarball with the YAML version.

1. Create manually the *submission.yaml* and all the *data_X.yaml* files. For each *data_X.yaml* file write line-by-line according to the equivalence:

data_file	
New "yaml" format	**Old "input" format**
name: dataX.yaml	
independent_variables:	
- header: {name: ‘bbbbbbbbbbbbbb‘}	*yheader
dependent_variables:	
- header: {name: ‘aaaaaaaaaaaaaaaaa‘}	*xheader:
qualifiers:	*qual:
values:	
- value: 1364.2 errors: - {symerror: 4.7, label: stat} - {symerror: 97, label: sys}	1364.2 +- 4.7 (DSYS=97)

submission.yaml	
New "yaml" format	**Old "input" format**
name:	
location	*location:
description	*dscomment:
keywords:	
- {name: reactions, ………….}	*reackey:
- {name: observables, ………}	*obskey:
- {name: cmenergies, ……….}	

2. Test *data_X.yaml* separately in the Sandbox (hepdata.net) or in a tarball with *submission.yaml* case-by-case.