



Graeme Watt (IPPP Durham)

12th Meeting of the Management Board for the Development,
Application and Validation of Nuclear Data and Codes (MBDAV)

Nuclear Energy Agency, 31st May 2023

<https://hepdata.net>

Email: info@hepdata.net

Forum: hepdata-forum.cern.ch

 Follow @HEPData

Code: <https://github.com/HEPData>

What is HEPData?

- Unique *open-access* repository for high-level **data** from almost 10,000 **H**igh-**E**nergy **P**hysics publications.
- HEP is an early adopter of Open Science (e.g. Web, arXiv).
- *Complementary* to other HEP information providers, e.g. INSPIRE-HEP (literature), PDG (particle properties), CERN Open Data (event-level data), Zenodo (files).
- **FAIR** data: **F**indable, **A**ccessible, **I**nteroperable, **R**eusable.
- Funded by UK **S**cience & **T**echnology **F**acilities **C**ouncil.
- Based in **I**nstitute for **P**article **P**hysics **P**henomenology (**IPPP**) at Durham University (UK), going back to 1970s.

HEPData redevelopments

hepdata.cedar.ac.uk (switched off in 2022)

- Upgrade 2005-2009 by Andy Buckley and Mike Whalley:
“*HepData reloaded: reinventing the HEP data archive*”,
PoS ACAT2010 (2010) 067 [[arXiv:1006.0517](#)].
- Relational database (MySQL) and Java web interface.

hepdata.net (new production site since 2017)

- Partnership with CERN Scientific Information Service.
- Complete rewrite in 2015-2016 by Eamonn Maguire.
- J. Phys.: Conf. Ser. 898 102006 [[arXiv:1704.05473](#)]

HEPData technologies

- Invenio v3 digital library framework (used by Zenodo).
 - Uses Python 3 and Flask micro web framework.
 - More recent InvenioRDM for turn-key repositories.
- Custom visualisation (using D3.js) and submission code.
- New input text format for data & metadata using YAML.
- New data model defined in a PostgreSQL database.
- Metadata indexed with OpenSearch for fast searching.
- Deployment at CERN using Docker/Kubernetes (2020-).

Code: <https://github.com/HEPData>

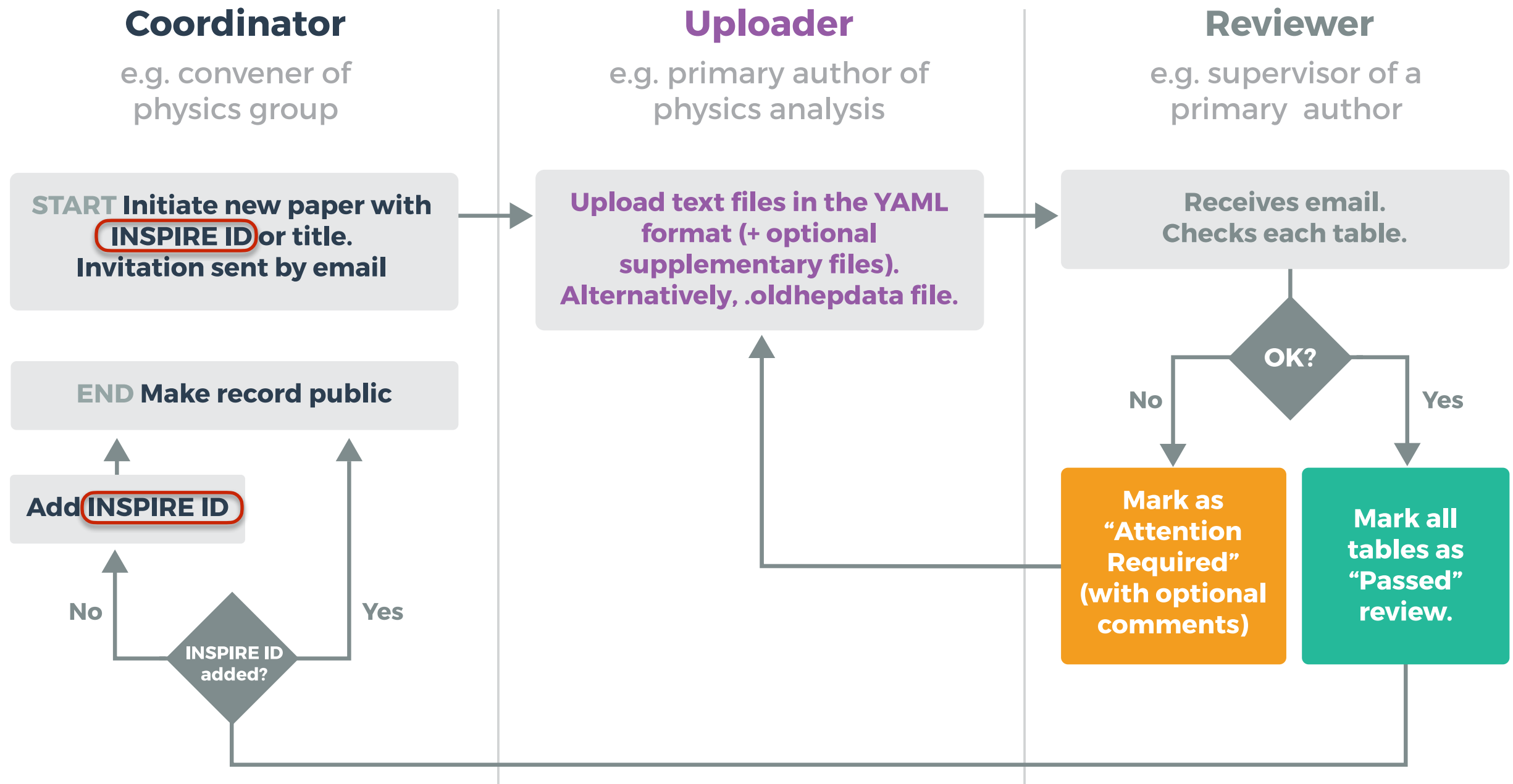
Modes of data entry

1. Manually harvested from data given in publications.
HEPData staff extracted tables from `.tex` source.
2. Data points directly submitted by experiments.
 - Pre-2014: no guidelines on preferred format.
 - Early 2014: encourage standard “input” format.
 - Late 2014: introduce online submission system.
 - Early 2017: allow submissions from hepdata.net.

Mode 1. now phased out in favour of mode 2.

Submission system on hepdata.net

<https://hepdata.net/submission>



- Submissions managed by Coordinators within each experiment/group.

Record from hepdata.net

HEPData

hepdata.net/record/ins1860766

HEPData

Search HEPData

Search

About

Submission Help

File Formats

Sign in

Browse all

Abudinén, F. et al.

Last updated on 2022-08-29 13:50

Accessed 278 times

Cite

JSON

Hide Publication Information

Search for $B^+ \rightarrow K^+ \nu \bar{\nu}$ Decays Using an Inclusive Tagging Method at Belle II

The Belle-II collaboration

Abudinén, F. , Adachi, I. , Adamczyk, K. , Ahlburg, P. , Aihara, H. , Akopov, N. , Aloisio, A. , Ky, N. Anh , Asner, D.M. , Atmacan, H.

Phys.Rev.Lett. 127 (2021) 181802, 2021.

<https://doi.org/10.17182/hepdata.130199>

Journal

INSPIRE

Resources

Abstract (data abstract)

SuperKEKB Belle II. Measurement of the branching fraction of $B^+ \rightarrow K^+ \nu \bar{\nu}$ at the Belle II experiment at the SuperKEKB. The analysed data sample corresponds to an integrated luminosity of 63 fb^{-1} collected at the $\Upsilon(4S)$ resonance and a sample of 9 fb^{-1} collected at an energy 60 MeV below the resonance between 2019-2021. Since no significant signal was observed, limit of 4.1×10^{-5} was set using CL_s method.

Download All

YAML with resource files

YAML

YODA

ROOT

CSV

corresponding to an...

Postfit yields off-resonance

Figure 1 in <https://journals.aps.org/prl/supplemental/10.17182/hepdata.130199.v1/t3>

Yields in off-resonance data and as predicted by the simultaneous fit to the on- and off-resonance data, corresponding to an...

Expected and observed limit

Figure 2 in <https://journals.aps.org/prl/supplemental/10.17182/hepdata.130199.v1/t4>

CL_s value as a function of the branching fraction of $B^+ \rightarrow K^+ \nu \bar{\nu}$ for expected and observed signal...

Selection efficiency

Figure 3 in <https://journals.aps.org/prl/supplemental/10.17182/hepdata.130199.v1/t5>

Signal efficiency as a function of the

Selection efficiency

[10.17182/hepdata.130199.v1/t5](https://doi.org/10.17182/hepdata.130199.v1/t5)

Resources

<https://www.hepdata.net>

Figure 3 in https://journals.aps.org/prl/supplemental/10.1103/PhysRevLett.127.181802/suppl_mat.pdf

Signal efficiency as a function of the dineutrino invariant mass squared q^2 for events in the signal region (SR) ($\text{BDT}_1 > 0.9$ and $\text{BDT}_2 > 0.95$). The error bars indicate the statistical uncertainty.

phrases

FCNC

$b \rightarrow s \ell \ell$ transition

electroweak penguin decay

missing energy

reactions

$B^+ \rightarrow K^+ \nu \bar{\nu}$

Luminosity

$63+9 \text{ fb}^{-1}$

$q^2 [\text{GeV}^2/c^4]$

Efficiency

0.0 - 2.0

12.66745696 ± 0.27207295

2.0 - 4.0

10.82571692 ± 0.26463688

4.0 - 6.0

7.04488885 ± 0.2278063

6.0 - 8.0

3.51769225 ± 0.1711566

8.0 - 10.0

1.46683133 ± 0.11813559

10.0 - 12.0

0.68175914 ± 0.08670158

12.0 - 14.0

0.27954132 ± 0.06074916

Visualize

Efficiency [%]

$q^2 [\text{GeV}^2/c^4]$

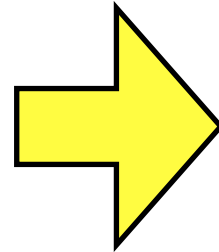
Belle II Simulation

- DOIs minted for whole (versioned) record and each table.

Data output formats

hepdata.net/formats

YAML: native
HEPData format.



submission.yaml
+ YAML data files for each table
+ optional resource files

- JSON: JavaScript Object Notation.
- CSV: comma-separated values.
- YODA: for inclusion in a Rivet analysis.
- ROOT: binary .root file.

<https://www.hepdata.net/record/ins1860766?format=csv>

- Programmatic access to download data in different formats.
- `format={original,yaml,json,csv,yoda,root}`
- Additional (optional) arguments for `version` and `table`.
-  `curl record w/o format` returns Schema.org JSON-LD.

hepdata-cli

- CLI and Python API for HEPData search/download/upload.
- Summer project in 2020 by Giuseppe De Laurentis.
- Install (in venv) with: `pip install hepdata-cli`
- Examples of usage:

```
hepdata-cli find 'collaborations:"Belle-II"' -i inspire
```

```
hepdata-cli fetch-names 1860766 -i inspire
```

```
hepdata-cli download 1860766 -f csv -i inspire
```

```
hepdata-cli upload /path/to/TestHEPSubmission.tar.gz -e  
my@email.com -p $PASSWORD -r 123456 -i $INVITATION_COOKIE -s False
```

Code: <https://github.com/HEPData/hepdata-cli>

submission.yaml file

- Separate YAML “document” for each data table, e.g.

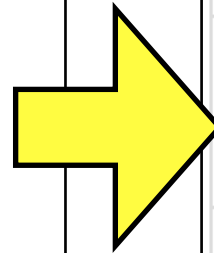
```
---
# This is Table 3.
name: "Table 3"
location: Data from Figure 8A
description: Normalized ZZ fiducial cross section (multiplied by 10^6 for
keywords: # used for searching, possibly multiple values for each keyword
- {name: reactions, values: [P P --> Z0 Z0 X]}
- {name: observables, values: [DSIG/DPT]}
- {name: cmenergies, values: [7000.0]}
- {name: phrases, values: [Inclusive, Single Differential Cross Section,
data_file: data3.yaml
additional_resources:
- {description: Image file, location: figFigure8A.png}
- {description: Thumbnail image file, location: thumb_figFigure8A.png}
```

- Validated against submission_schema.json

YAML data file

- “Independent” and “dependent” variables:

```
independent_variables:
- header: {name: Leading dilepton PT, units: GEV}
  values:
  - {low: 0, high: 60}
  - {low: 60, high: 100}
  - {low: 100, high: 200}
  - {low: 200, high: 600}
dependent_variables:
- header: {name: 10**6 * 1/SIG(fiducial) * D(SIG(fiducial))/DPT, units: GEV**-1}
  qualifiers:
  - {name: RE, value: P P --> Z0 < LEPTON+ LEPTON- > Z0 < LEPTON+ LEPTON- > X}
  - {name: SQRT(S), units: GEV, value: 7000}
  values:
  - value: 7000
    errors:
    - {symerror: 1100, label: stat}
    - {symerror: 79, label: 'sys,detector'}
    - {symerror: 15, label: 'sys,background'}
  - value: 9800
    errors:
    - {symerror: 1600, label: stat}
    - {symerror: 75, label: 'sys,detector'}
    - {symerror: 15, label: 'sys,background'}
  - value: 1600
    errors:
    - {symerror: 490, label: stat}
    - {symerror: 41, label: 'sys,detector'}
    - {symerror: 2, label: 'sys,background'}
  - value: 80
    errors:
    - {symerror: 60, label: stat}
    - {symerror: 2, label: 'sys,detector'}
    - {symerror: 0, label: 'sys,background'}
```



RE	P P --> Z0 < LEPTON+ LEPTON- > Z0 < LEPTON+ LEPTON- > X
SQRT(S)	7000.0 GeV
Leading dilepton PT [GEV]	$10^{**6} * 1/SIG(fiducial) * D(SIG(fiducial))/DPT [GEV^{**-1}]$
0.0 - 60.0	7000.0 ± 1100.0 stat ± 79.0 sys,detector ± 15.0 sys,background
60.0 - 100.0	9800.0 ± 1600.0 stat ± 75.0 sys,detector ± 15.0 sys,background
100.0 - 200.0	1600.0 ± 490.0 stat ± 41.0 sys,detector ± 2.0 sys,background
200.0 - 600.0	80.0 ± 60.0 stat ± 2.0 sys,detector

- Validated against data_schema.json

Covariance matrices

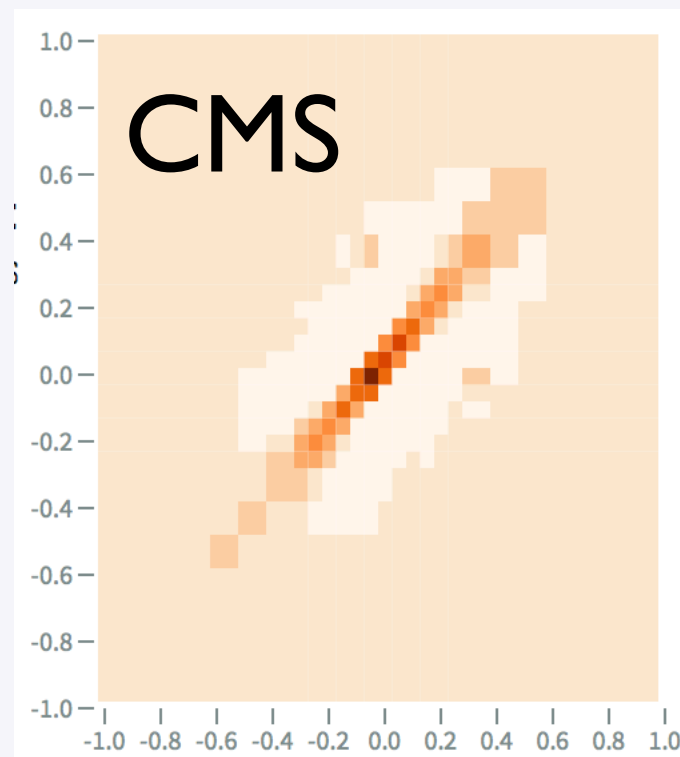
Correlation/covariance matrices

hepdata-submission.readthedocs.io

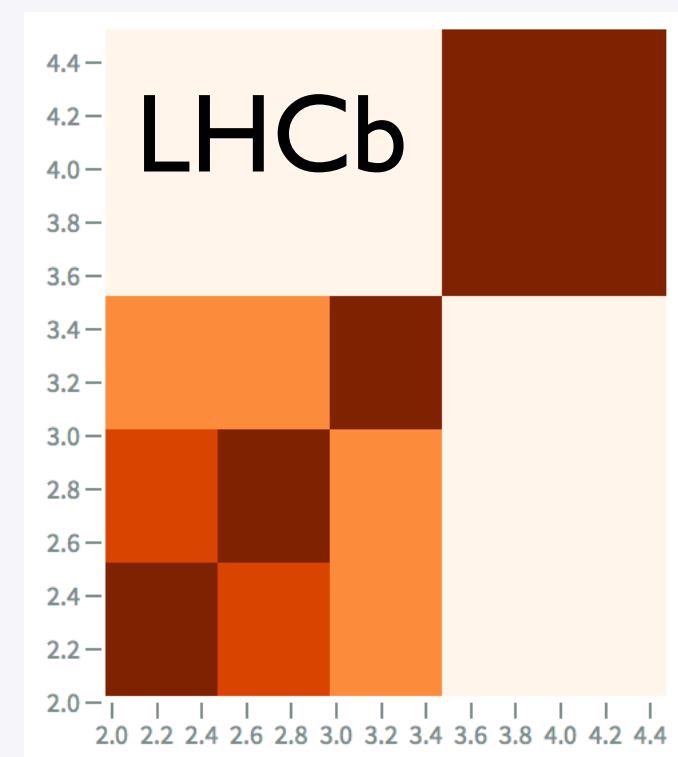
Correlation/covariance matrices can be encoded in a format with two independent variables (giving the bins) and one dependent variable (giving the covariance/correlation), e.g.

```
independent_variables:
- header: {name: PTjet, units: GeV}
  values:
  - {low: 25, high: 45}
  - {low: 45, high: 65}
  - {low: 45, high: 65}
  ...
- header: {name: PTjet, units: GeV}
  values:
  - {low: 25, high: 45}
  - {low: 25, high: 45}
  - {low: 45, high: 65}
  ...
dependent_variables:
- header: {name: Correlation}
  values:
  - {value: 1.0000}
  - {value: 0.8727}
  - {value: 1.0000}
  ...
```

ins | 605749



ins | 454404



- Export to **JSON**, **CSV**, **ROOT** (**TGraph2DErrors**, **TH2F**), **YODA** (**Scatter3D**).

Additional resource files

The screenshot shows the HEPData website interface. A modal window titled 'Additional Publication Resources' is open, displaying a list of resources on the left and their details on the right. The left sidebar lists 'Common Resources' (8 items) including systematic tables for SR_Gtt_0L_B, SR_Gtt_0L_M1, SR_Gtt_0L_M2, SR_Gtt_0L_C, SR_Gtt_1L_B, SR_Gtt_1L_M1, SR_Gtt_1L_M2, SR_Gtt_1L_C, and SR_Gbb_B. The right panel shows details for two 'dat File' resources and two 'External Link' resources. The 'dat File' resources are param cards for Gbb 2000 and Gtb 2200 models. The 'External Link' resources are a webpage with all figures and tables, and an arXiv link. The background shows the main record page for 'Search for supersymmetry in final states with missing transverse momentum in three or more b -jets in 139 fb⁻¹ of proton-proton collisions at $\sqrt{s}=13$ TeV with the ATLAS detector'.

HEPData

Search for supersymmetry in final states with missing transverse momentum in three or more b -jets in 139 fb⁻¹ of proton-proton collisions at $\sqrt{s}=13$ TeV with the ATLAS detector

The ATLAS collaboration

CERN-EP-2022-213, 2022.

<https://doi.org/10.17182/hepdata.95928>

INSPIRE Resources

HistFactory

Abstract (data abstract)

A search for supersymmetry involving the pair production of gluinos decaying via off-shell third-generation squarks. The lightest neutralino ($\tilde{\chi}_1^0$) is reported. It exploits proton-proton collision data at a centre-of-mass energy of 13 TeV with an integrated luminosity of 139 fb⁻¹ collected by the ATLAS detector from 2015 to 2018. The search targets final states containing large missing transverse momentum, an electron or muon, and several energetic jets, at least one of which must be identified as containing b -hadrons. Simple kinematic event selection and an event selection based upon a deep neural-network are used. No significant signal is observed. The 95% CL upper limit on the product of the gluino production cross-section and branching ratio to $b\bar{b}\mu^+\mu^-$ is 1.1 fb.

Additional Publication Resources

filter

Common Resources 8

Systematic table for SR_Gtt_0L_B 2

Systematic table for SR_Gtt_0L_M1 2

Systematic table for SR_Gtt_0L_M2 2

Systematic table for SR_Gtt_0L_C 2

Systematic table for SR_Gtt_1L_B 2

Systematic table for SR_Gtt_1L_M1 2

Systematic table for SR_Gtt_1L_M2 2

Systematic table for SR_Gtt_1L_C 2

Systematic table for SR_Gbb_B 2

dat File

description: Param card (SLHA file) for Gbb 2000, 1000 model, location: param_card_376017.dat

[10.17182/hepdata.95928.v1/r2](https://doi.org/10.17182/hepdata.95928.v1/r2)

Download

dat File

description: Param card (SLHA file) for Gtb 2200, 600 model, location: param_card_376093.dat

[10.17182/hepdata.95928.v1/r3](https://doi.org/10.17182/hepdata.95928.v1/r3)

Download

C++ File

description: Code for NN and CC regions in SimpleAnalysis, location: ANA-SUSY-2018-30.cxx

[10.17182/hepdata.95928.v1/r4](https://doi.org/10.17182/hepdata.95928.v1/r4)

Download

HistFactory File

Archive of full likelihoods in the HistFactory JSON format

[10.17182/hepdata.95928.v1/r5](https://doi.org/10.17182/hepdata.95928.v1/r5)

Download

External Link

Webpage with all figures and tables

[View Resource](#)

External Link

arXiv

[View Resource](#)

ttbar normalisation 4.9

15

10

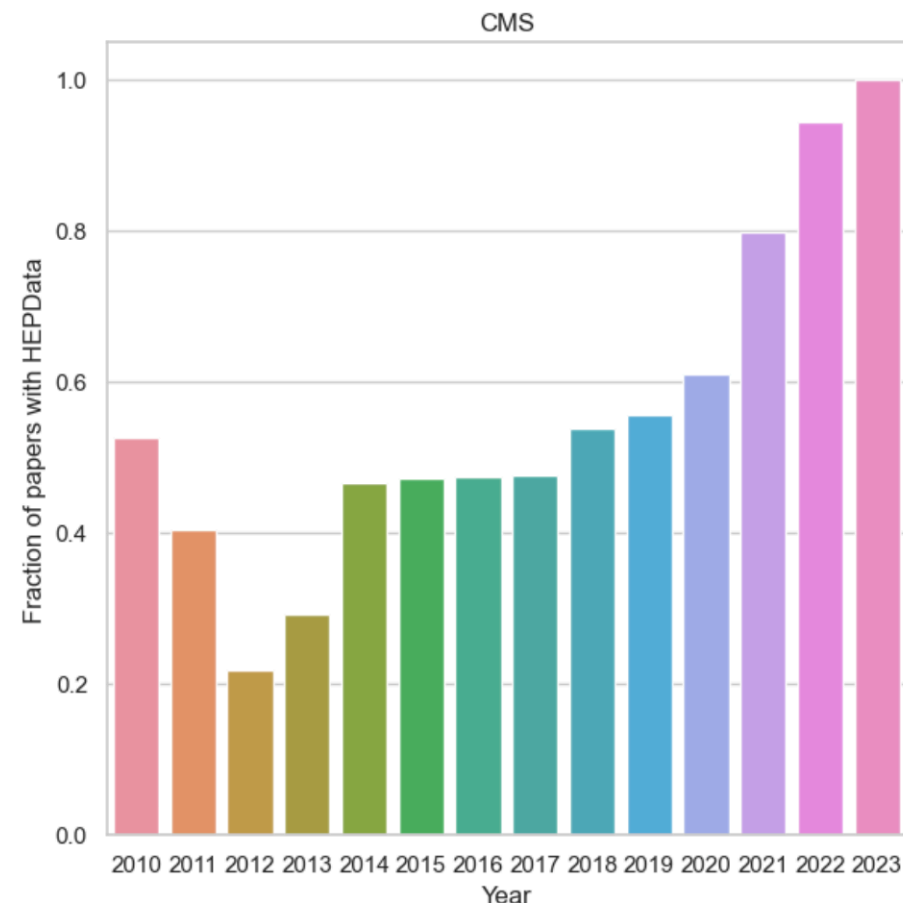
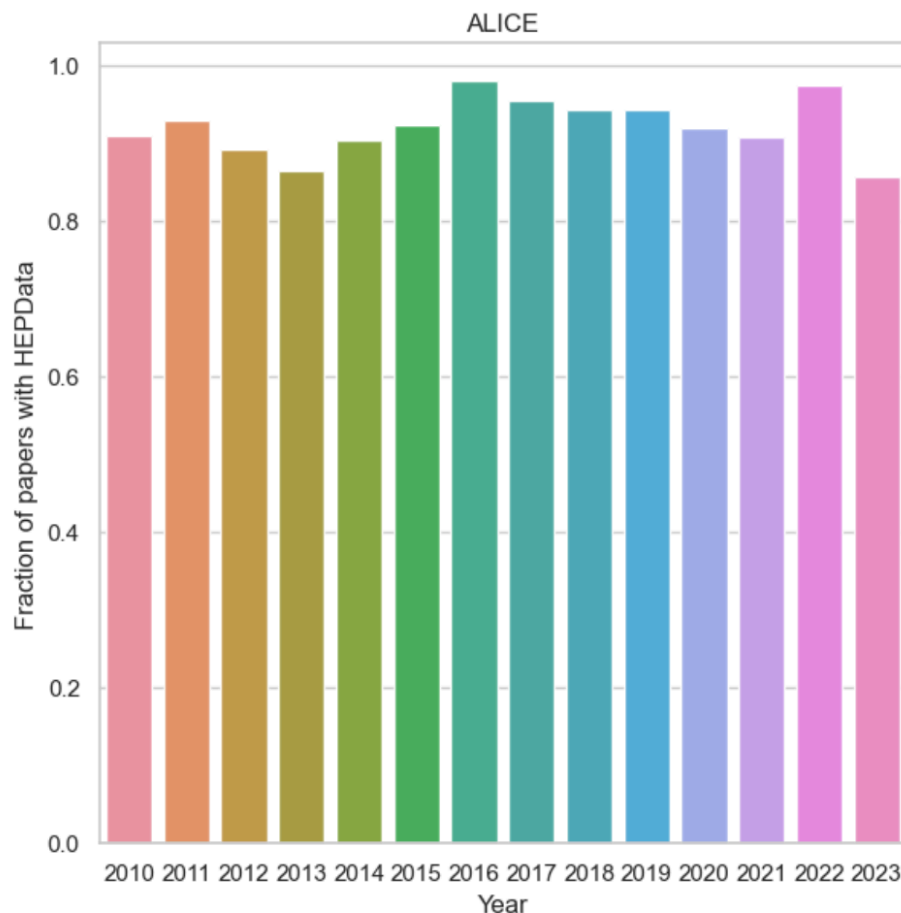
5

Total Uncertainty Experimental Theoretical MC statistical ttbar normalisation

- DOIs now minted via DataCite for additional resource files.

Effect of policies

- CERN Open Data Policy for the LHC Experiments (2020):
“To maximise the scientific value of their publications, the experiments will make public additional information and data at the time of publication, stored in collaboration with portals such as HEPData”
- Fraction of LHC publications with HEPData records.
ALICE: HEPData is prerequisite, CMS: mandate from 2020.

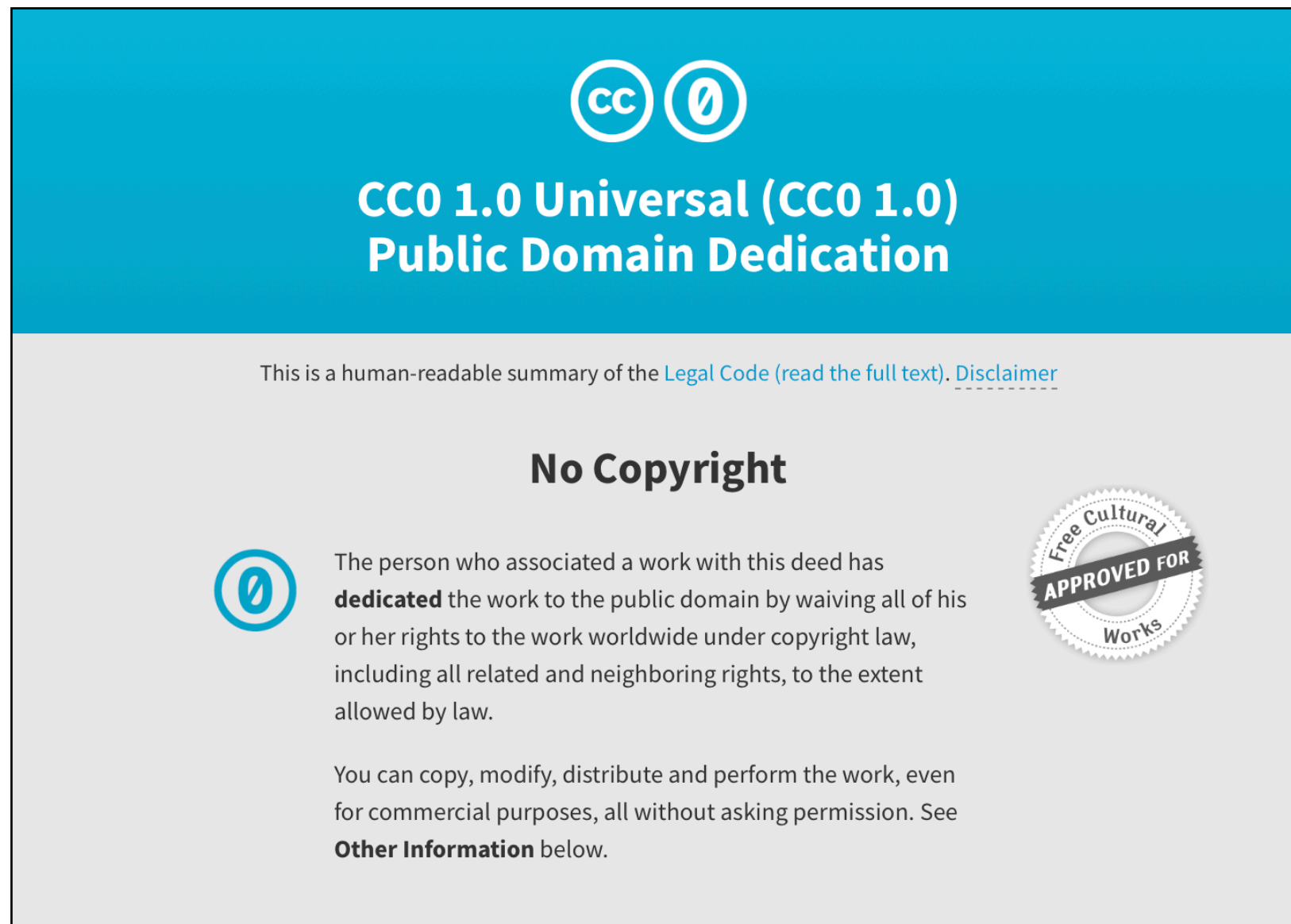


Plots:
[GitHub](#)
[Binder](#)

Terms of Use

<https://www.hepdata.net/terms>

- “Unless specified otherwise for selected datasets, all metadata and datasets in the HEPData service are made available under the terms of CC0”



The image is a screenshot of the Creative Commons CC0 1.0 Universal Public Domain Dedication license page. The top section has a blue background with the CC0 logo (two circles, one with 'cc' and one with '0') and the text 'CC0 1.0 Universal (CC0 1.0) Public Domain Dedication'. Below this, on a light gray background, is a line of text: 'This is a human-readable summary of the [Legal Code \(read the full text\)](#). [Disclaimer](#)'. The main heading is 'No Copyright'. To the left of the text is a blue circle with a white '0'. The text explains that the person who associated a work with this deed has **dedicated** the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. It also states that you can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See **Other Information** below. On the right side, there is a circular seal that says 'Free Cultural Works' around the edge and 'APPROVED FOR' in the center.


**CC0 1.0 Universal (CC0 1.0)
Public Domain Dedication**

This is a human-readable summary of the [Legal Code \(read the full text\)](#). [Disclaimer](#)

No Copyright

 The person who associated a work with this deed has **dedicated** the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See **Other Information** below.



Summary

Email: info@hepdata.net

Forum: hepdata-forum.cern.ch

- **HEPData** is *the* repository for publication-related HEP data.
- Transition in 2017 to modern hepdata.net site hosted by CERN.
- Responsibility for data submission now held by experiments.
- Native YAML format with visualisation and format conversion.
- [hepdata-cli](https://github.com/HEPData/hepdata-cli) search/download/upload from command line.
- Policies within experiments/laboratories affect submission rate.
- HEPData has “[Terms of Use](#)” for data with implicit [CC0](#) license.
- Interested in perspectives from the nuclear data community.

Code: <https://github.com/HEPData>