



Graeme Watt (IPPP Durham)

Belle II Data Preservation Workshop, 7th October 2022

<https://hepdata.net>

Email: info@hepdata.net

Forum: hepdata-forum.cern.ch

 Follow @HEPData

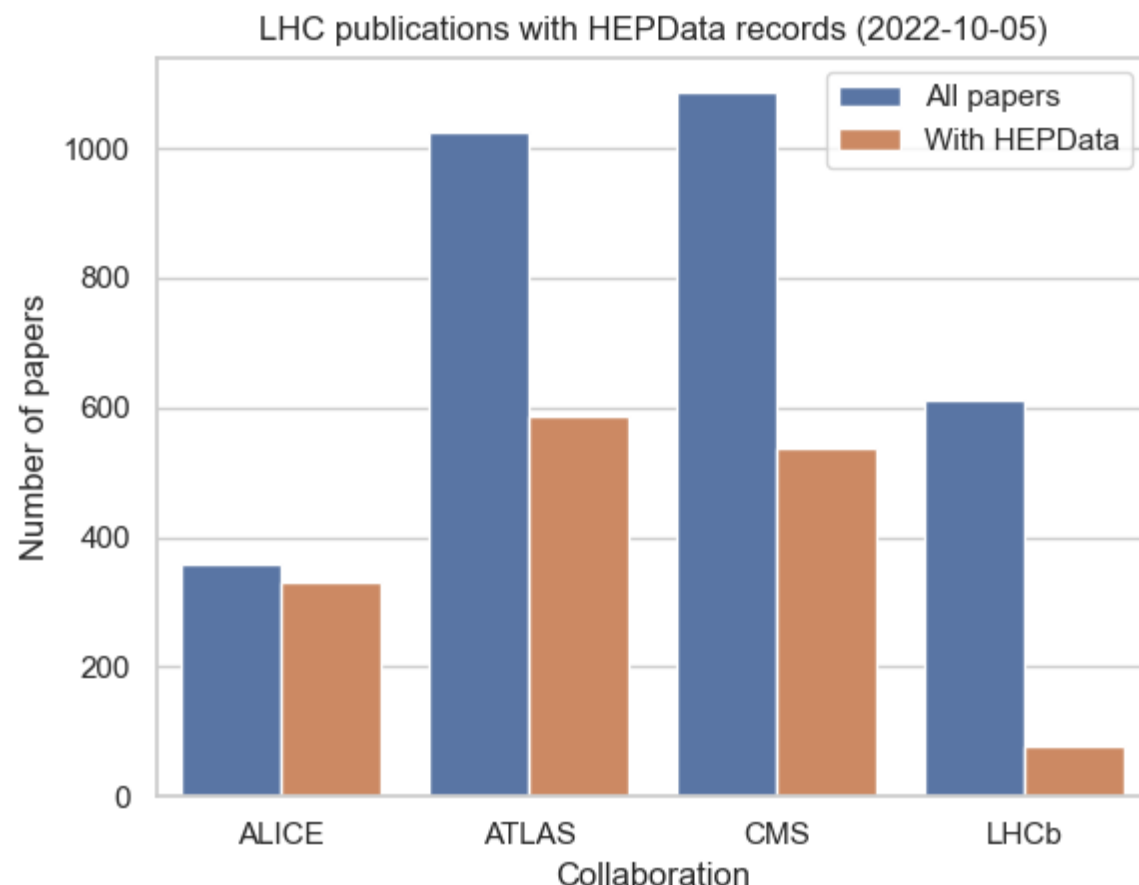
Code: <https://github.com/HEPData>

What is HEPData?

- Unique *open-access* repository for high-level **data** from almost 10,000 **HEP** (“hep-ex” or “nucl-ex”) papers.
- **FAIR** data: **F**indable, **A**ccessible, **I**nteroperable, **R**eusable.
- *Complementary* to other HEP information providers, e.g. INSPIRE-HEP (literature), PDG (particle properties), CERN Open Data (event-level data), Zenodo (files).
- Based in **I**nstitute for **P**article **P**hysics **P**henomenology (**IPPP**) at Durham University (UK), going back to 1970s.
- Transition in 2017 to hepdata.net site, hosted at CERN. Partnership with CERN Scientific Information Service.
J. Phys.: Conf. Ser. 898 102006 [arXiv:1704.05473]

Search from inspirehep.net

- New INSPIRE (blog) shares many features with HEPData.
- INSPIRE publication records link to hepdata.net records.
- Search INSPIRE for publication records with HEPData.
- LHC publications with HEPData (Jupyter Notebook):



INSPIRE search query:

- hep-ex or nucl-ex
- Published in a journal
- Not conference paper

ALICE: **92%**
ATLAS: **57%**
CMS: **49%**
LHCb: **13%**

Belle: $31/533 = 6\%$
Belle-II: $1/9 = 11\%$

Search from hepdata.net

The screenshot shows the HEPData website interface. The search bar at the top contains the query 'collaborations:Belle', which is highlighted with a red box. Below the search bar, there are filters for 'Max results', 'Sort by', and 'Reverse order'. The search results show 10 of 39 results. On the left side, there is a 'Collaboration' filter with a red box around it, showing 'BELLE' with 38 results and 'BELLE-II' with 1 result. Below this is a 'Subject_areas' filter with 'hep-ex' (38), 'Phenomenology-HEP' (2), 'hep-ph' (2), and 'HEP Experiment' (1). At the bottom left, there is a 'Phrases' filter with 'Exclusive' (30). The main content area displays the first search result: 'Measurement of Differential Branching Fractions of Inclusive $B \rightarrow X_u \ell^+ \nu_\ell$ Decays' by The Belle collaboration. The result includes the journal 'Phys.Rev.Lett. 127 (2021) 261801, 2021.' and a link to the Inspire Record 1895149. The abstract describes the first measurements of differential branching fractions of inclusive semileptonic $B \rightarrow X_u \ell^+ \nu_\ell$ decays. Below the abstract, there are three figures: Fig. 3 (top left) showing differential branching fractions as a function of lepton energy, Fig. 3 (top right) showing differential branching fractions as a function of four-momentum-transfer squared, and Fig. 3 (middle left) showing differential branching fractions as a function of invariant hadronic mass. A 'More...' link is also present. At the bottom of the page, there is a link to 'Search for $B^+ \rightarrow K^+ \nu \bar{\nu}$ Decays Using an Inclusive Tagging Method at Belle II'.

HEPData

hepdata.net/search/?q=collaborations%3ABelle&sort_by=latest

About Submission Help File Formats Sign in

collaborations:Belle Search Reset search Advanced JSON

Max results Sort by Reverse order Showing 10 of 39 results

Date

Collaboration

BELLE 38

BELLE-II 1

Subject_areas

hep-ex 38

Phenomenology-HEP 2

hep-ph 2

HEP Experiment 1

Phrases

Exclusive 30

Measurement of Differential Branching Fractions of Inclusive $B \rightarrow X_u \ell^+ \nu_\ell$ Decays

The Belle collaboration Cao, L. ; Sutcliffe, W. ; Van Tonder, R. ; *et al.*

Phys.Rev.Lett. 127 (2021) 261801, 2021.

Inspire Record 1895149 DOI 10.17182/hepdata.131599

The first measurements of differential branching fractions of inclusive semileptonic $B \rightarrow X_u \ell^+ \nu_\ell$ decays are performed using the full Belle data set of 711 fb^{-1} of integrated luminosity at the $\Upsilon(4S)$ resonance and for $\ell = e, \mu$. Differential branching fractions are reported as a function of the lepton momentum, the four-momentum-transfer squared, light-cone momenta, the hadronic mass, and the hadronic mass...

50 data tables match query

Fig. 3 (top left) The measured differential branching fractions as a function of the lepton energy in the B rest frame (E_ℓ^B).

Fig. 3 (top right) The measured differential branching fractions as a function of the four-momentum-transfer squared of the B to the X_u system q^2 .

Fig. 3 (middle left) The measured differential branching fractions as a function of the invariant hadronic mass of the X_u system (M_X).

More...

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

Connection to Rivet analyses

The screenshot shows the HEPData website interface. The browser address bar displays the URL: `hepdata.net/search/?q=collaborations%Belle+AND+analysis%3Arivet&sort_by=latest`. The search bar contains the query `collaborations:Belle AND analysis:rivet`, which is highlighted with a red box. The search results are displayed on the right side of the page, showing two entries. The first entry is titled "Precise determination of the CKM matrix element $|V_{cb}|$ with $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decays with hadronic tagging at Belle" and is associated with the Belle collaboration. The second entry is titled "Measurement of the decay $B \rightarrow D \ell \nu_\ell$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{cb}|$ " and is also associated with the Belle collaboration. Both entries are marked with a "Rivet Analysis" icon, which is also highlighted with a red box. On the left side of the page, there are filters for Date, Collaboration, Subject_areas, and Phrases. The Collaboration filter shows "BELLE" selected, which is highlighted with a red box. The Subject_areas filter shows "hep-ex" selected. The Phrases filter shows "Exclusive" selected.

- HEPData exports data in YODA format for Rivet.

Record from hepdata.net

HEPData

hepdata.net/record/ins1860766

Search HEPData

Search

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...

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

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.

Efficiency [%]

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

Belle II Simulation

phrases

- FCNC
- $b \rightarrow s \ell \ell$ transition
- electroweak penguin decay
- missing energy

reactions

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

Visualize

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

Efficiency [%]

Luminosity	63+9 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

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>

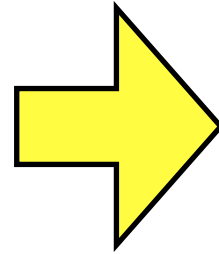
Signal efficiency as a function of the

- 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.

Modes of data entry

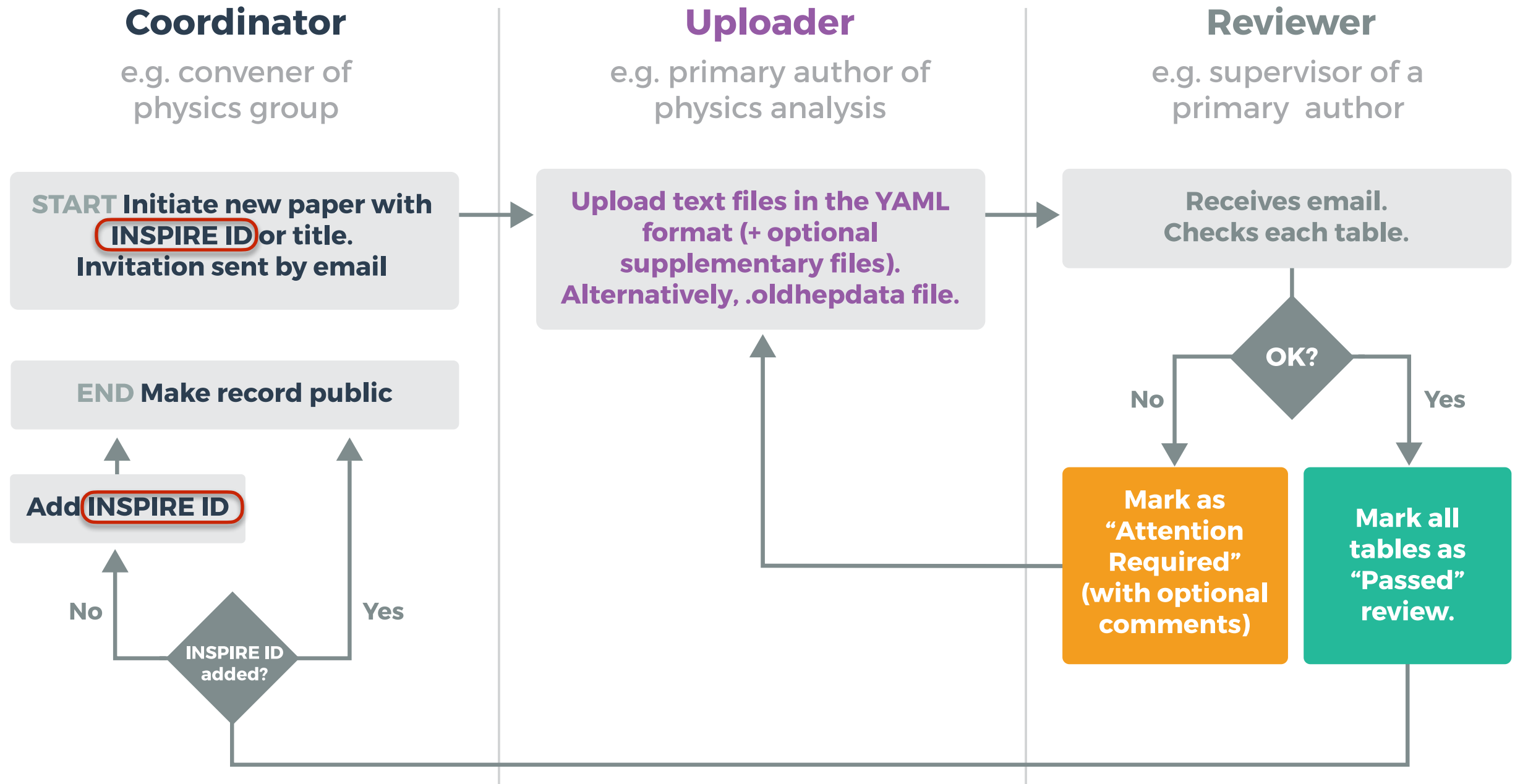
1. Manually harvested from “`hep-ex`” / “`nucl-ex`”.
HEPData staff extracted tables from `.tex` source.
2. Data points directly submitted by experiments.
 - Pre-2014: no guidelines on preferred format.
 - Early 2014: encourage standard “input” format.
 - Late 2014: introduce online submission system.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

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

Submission system on hepdata.net

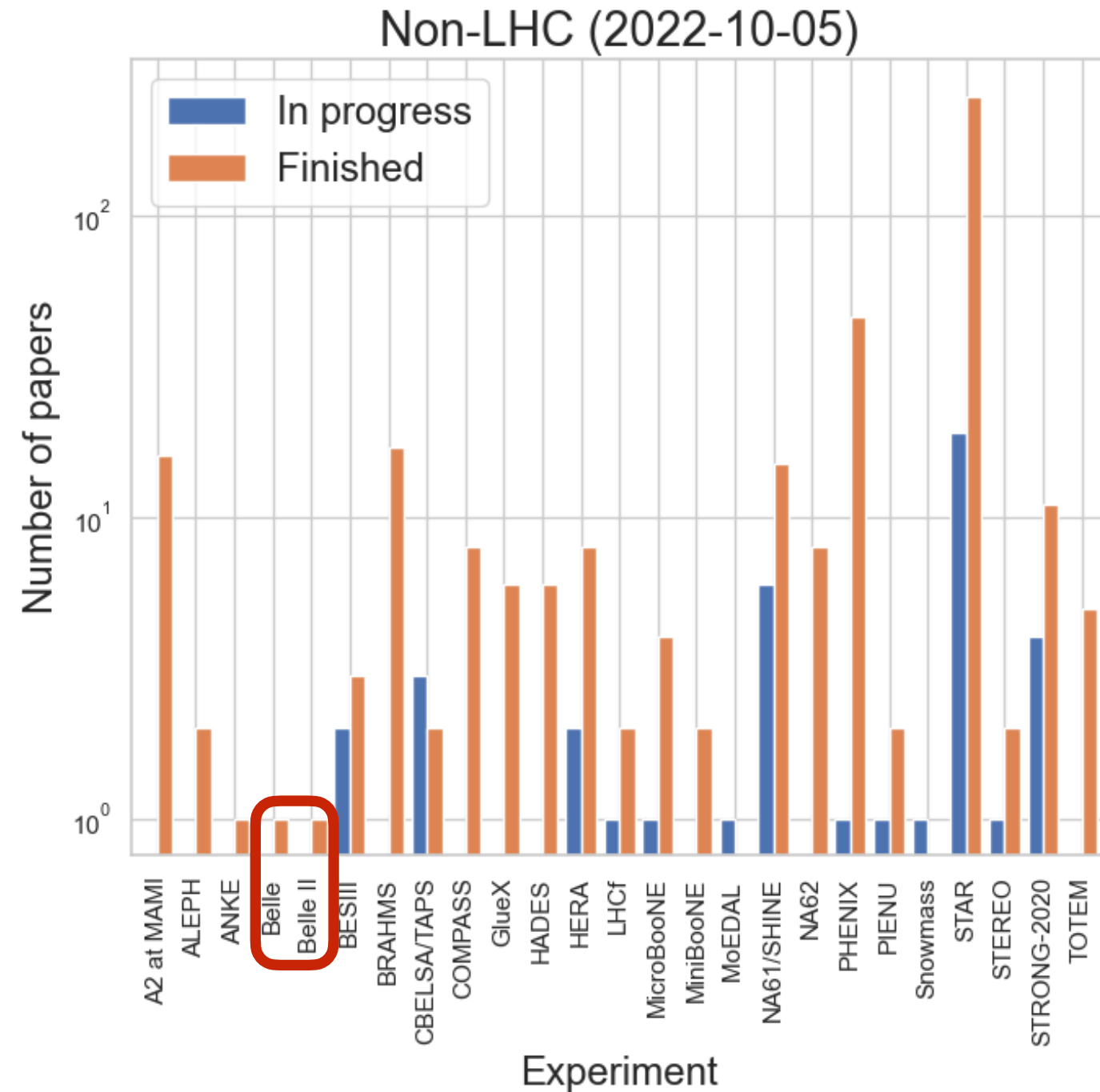
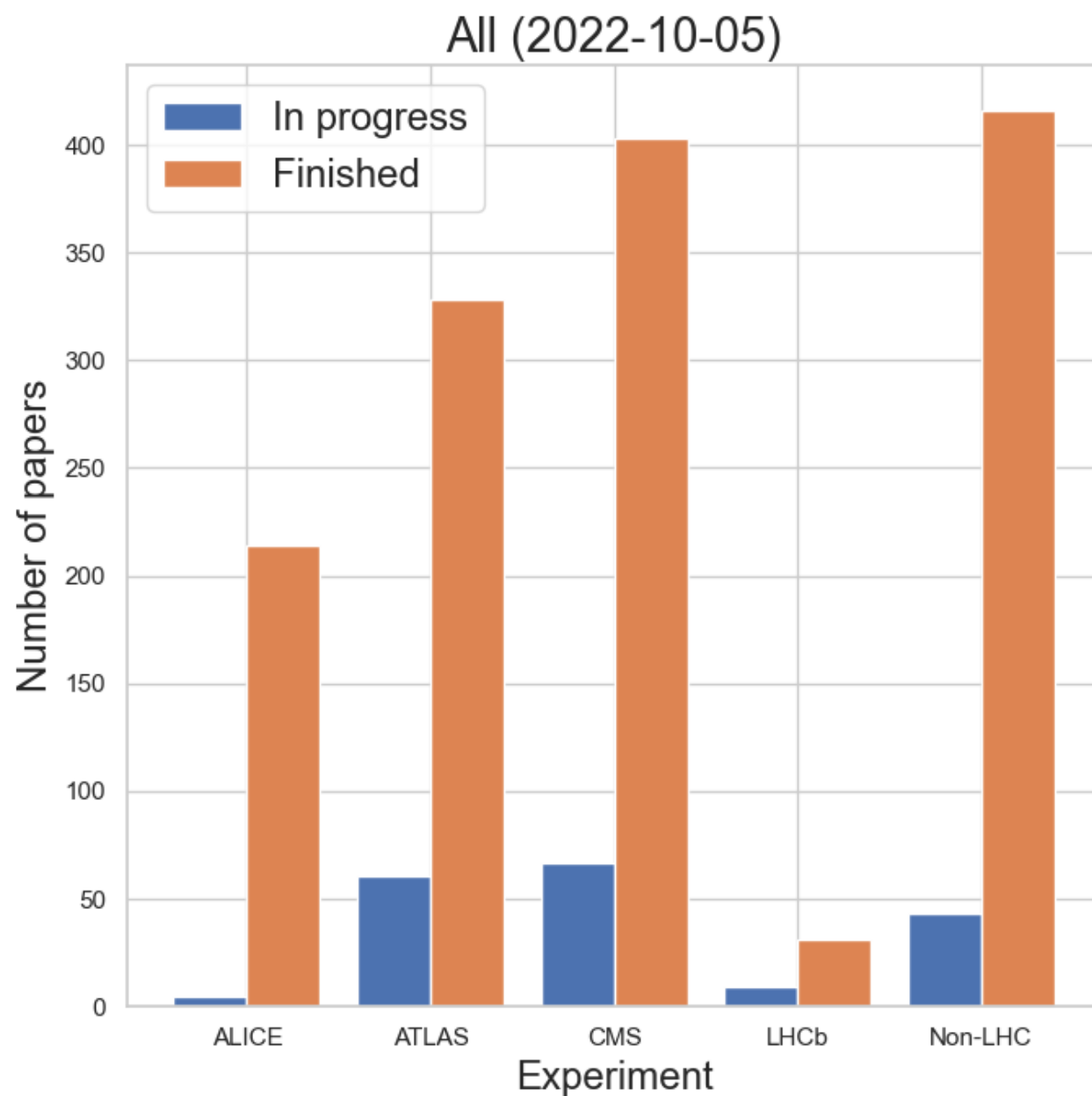
Belle mgt-hepdata-coordinators-belle@belle2.org
Belle II mgt-hepdata-coordinators@belle2.org

<https://hepdata.net/submission>



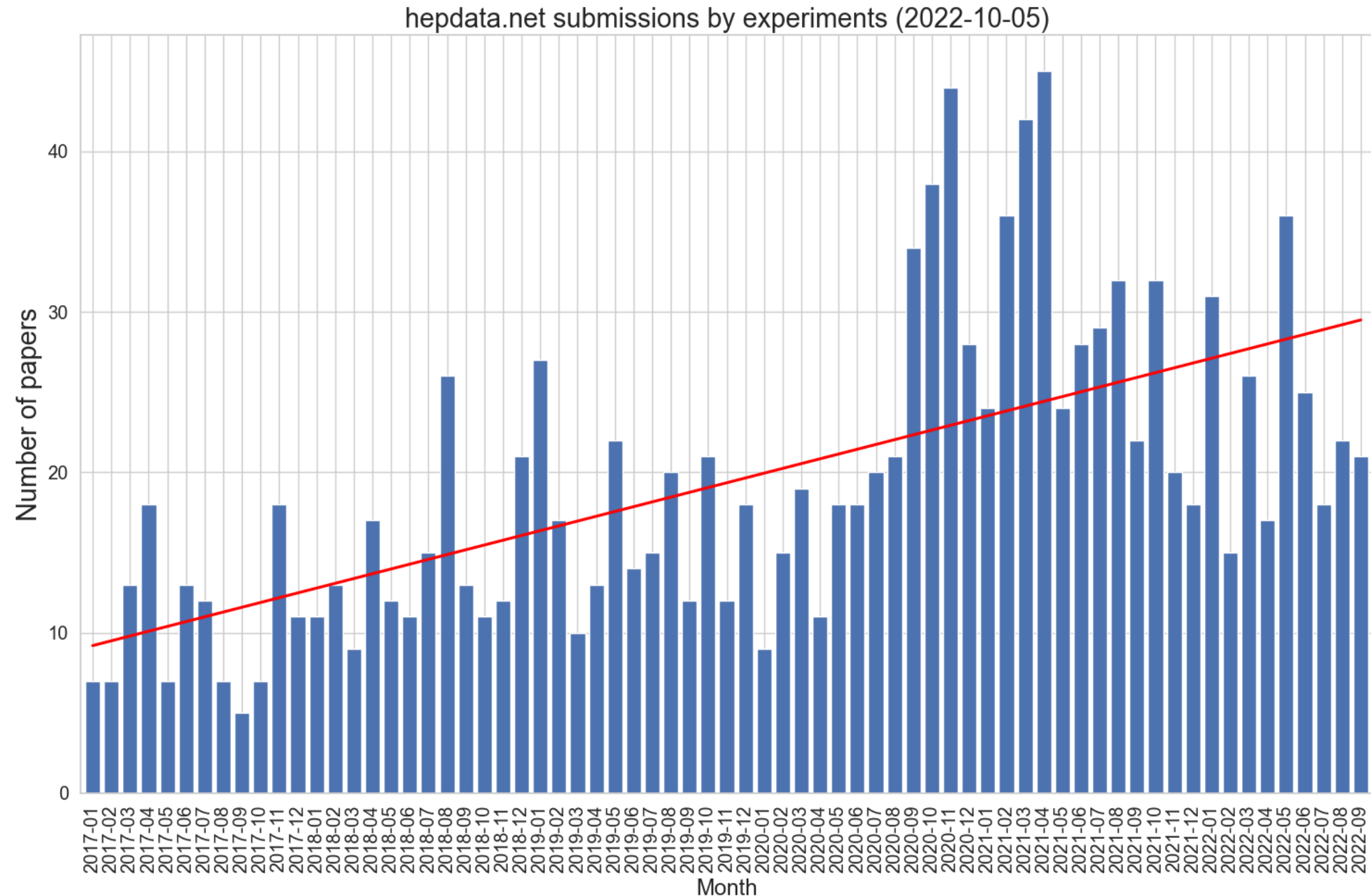
- Also a Sandbox for any user to test uploads without special privileges.

Submission system usage (01/2017-)



- Total of **1392** finished submissions, comprising **ALICE (214)**, **ATLAS (328)**, **CMS (403)**, **LHCb (31)**, **Non-LHC (416)**.

Submissions per month from 2017 to 2022



- Year-on-year growth: **125** in **2017**, **171** in **2018**, **201** in **2019**, **275** in **2020**, **352** in **2021**.

submission.yaml file

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

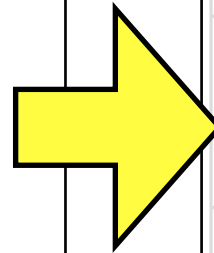
```
---
# This is Table 3.
name: "Table 3"
location: Data from Figure 8A
description: Normalized ZZ fiducial cross section (multiplied by 10^6 for
keywords: # used for searching, possibly multiple values for each keyword
- {name: reactions, values: [P P --> Z0 Z0 X]}
- {name: observables, values: [DSIG/DPT]}
- {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

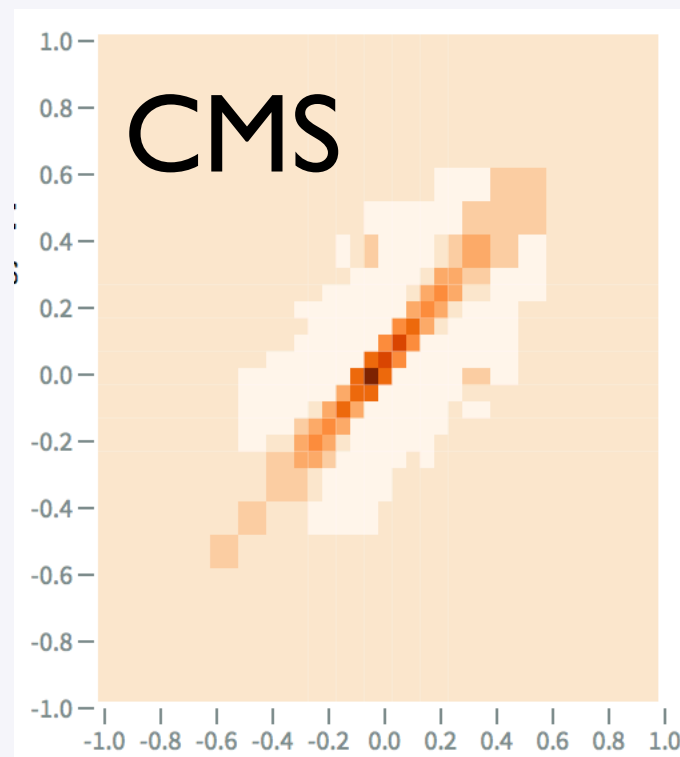
Correlation/covariance matrices

hepdata-submission.readthedocs.io

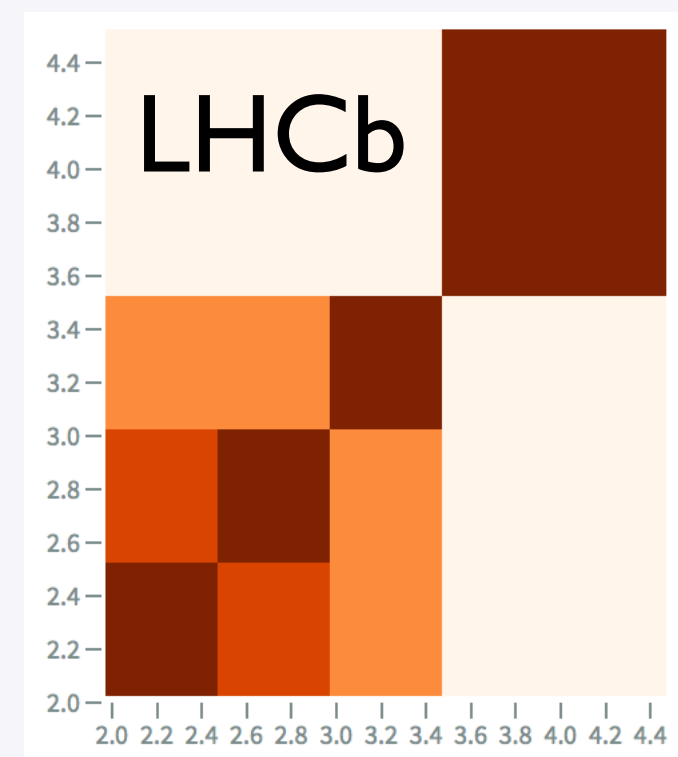
Correlation/covariance matrices can be encoded in a format with two independent variables (giving the bins) and one dependent variable (giving the covariance/correlation), e.g.

```
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**).

Submission documentation

- Documentation at hepdata-submission.readthedocs.io. Includes example Python scripts ([simple](#), [complicated](#)).
- HEPData YAML files checked against [JSON schema](#) by [validation code](#) during submission. [hepdata-validate](#)
- [hepdata_lib](#) package by *Clemens Lange* and *Andreas Albert*. Library to read in text/ROOT and write HEPData YAML. https://github.com/HEPData/hepdata_lib
- Similar Python package by *Christian Holm Christensen*. <https://gitlab.com/cholmcc/hepdata>
- Also [hepdata_maker](#) tool by *Krzysztof Bożek* (ATLAS).
- Experiments often develop internal HEPData tools/docs.

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>

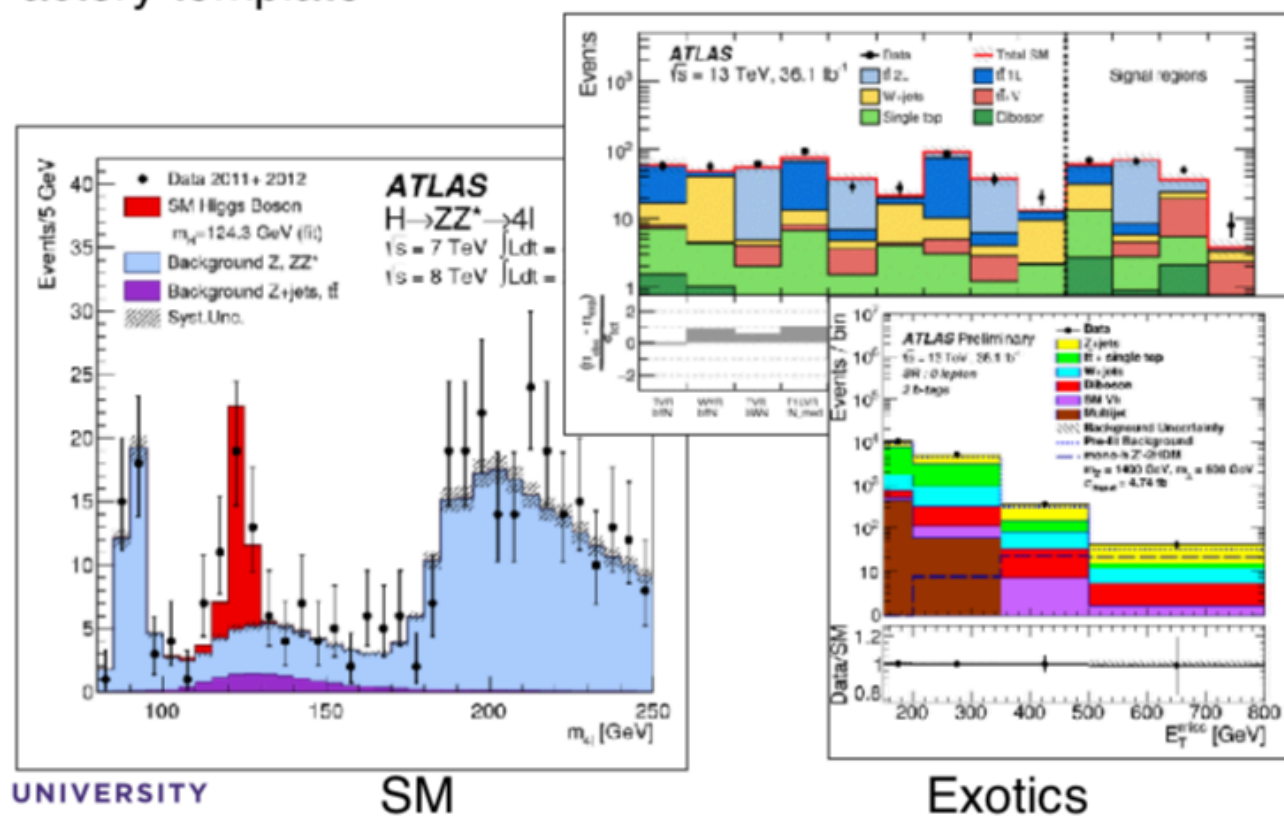
HistFactory

HistFactory

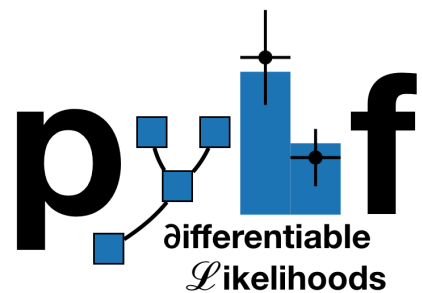
Lukas Heinrich

fundamentally a (quite flexible) p.d.f template to build **statistical models** from binned distributions and data.

~all binned, folded, frequentist statistical models in ATLAS are expressed using HistFactory template



- Based on simple **ROOT** histograms organised in an **XML** file.
- ROOT/XML workspace replaced by plain-text **pyhf JSON**.



pyhf likelihoods

See later talk by
Lukas Heinrich



The screenshot shows the HEPData website interface. The search bar at the top contains the query 'analysis:HistFactory'. The left sidebar shows filters for 'Collaboration' (ATLAS is selected), 'Subject_areas' (hep-ex), and 'Phrases' (Proton-Proton Scattering, SUSY, Supersymmetry, Electroweak, Limits). The main content area displays two search results, both marked with a 'HistFactory' icon. The first result is titled 'Search for chargino-neutralino pair production in final states with three leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector' and is from 'The ATLAS collaboration' (Aad, Georges; Abbott, Braden Keim; Abbott, Dale; et al.), published in 'Eur.Phys.J.C 81 (2021) 1118, 2021'. The second result is titled 'Searches for new phenomena in events with two leptons, jets, and missing transverse momentum in 139 fb⁻¹ of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector' and is also from 'The ATLAS collaboration', published in 'CERN-EP-2022-014, 2022'. Both results include links to Inspire Record and DOI.

- Separate DOIs now allocated for attached pyhf archive files.

Summary

Email: info@hepdata.net

Forum: hepdata-forum.cern.ch

- **HEPData** is *the* repository for publication-related HEP data.
- *Caveats:* design restricts size (~MB) and format (mostly tabular).
- Responsibility for data submission now held by experiments.
- Belle II Coordinators: mgt-hepdata-coordinators@belle2.org
-  A Belle and a Belle II record completed on 29th August 2022.
- Documentation and software tools available to assist upload.
- hepdata-cli search/download/upload from command line.
-  Enhanced linking to Rivet analyses and pyhf likelihood data.

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