



HEPData

status

Graeme Watt (Manager)

HEPData Users' Meeting 2025

IPPP Durham, 15th January 2025

<https://hepdata.net>

Email: info@hepdata.net

Forum: hepdata-forum.cern.ch

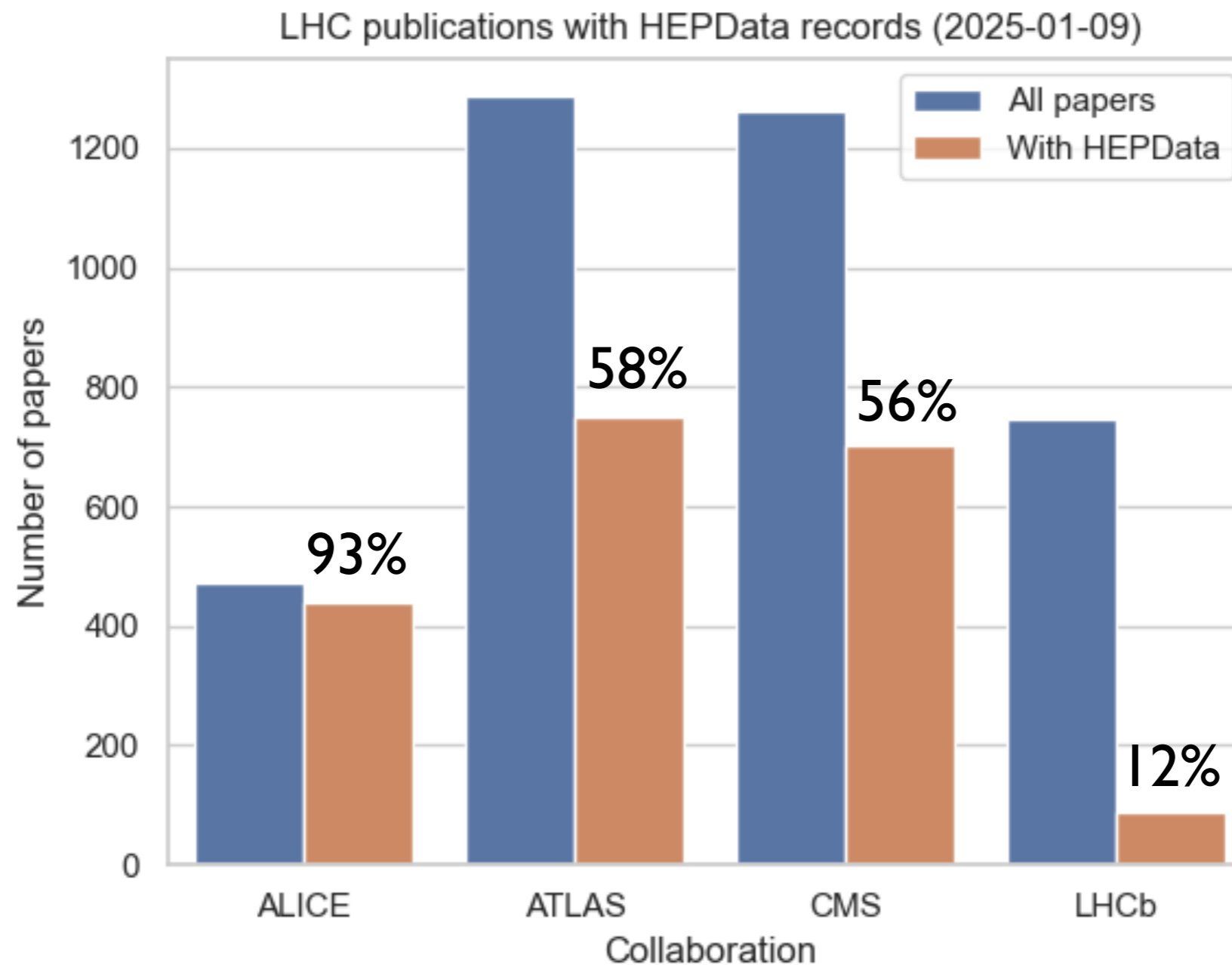
 Follow @HEPData

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

Introduction

- **This meeting:** continuation of previous HEPData advisory board meetings, last held in January 2020.
- HEPData funded by UK **Science & Technology Facilities Council (STFC)** grant to **Institute for Particle Physics Phenomenology (IPPP)** at Durham University (UK).
- *IPPP Computing meeting* in February 2024 (HEPData talk).
- Current STFC grant (**10/2022 - 09/2026**) funds two staff:
 1. Graeme Watt (management and user support, **10/2013 -**)
 2. Software developer from Advanced Research Computing:
 - Alison Clarke (**11/2019 - 05/2022**)
 - Zeynep Akı (**04/2022 - 05/2022**)
 - Jordan Byers (**06/2022 -**)
- Partnership with CERN Scientific Information Service for infrastructure and operational support (Benjamin Bergia).

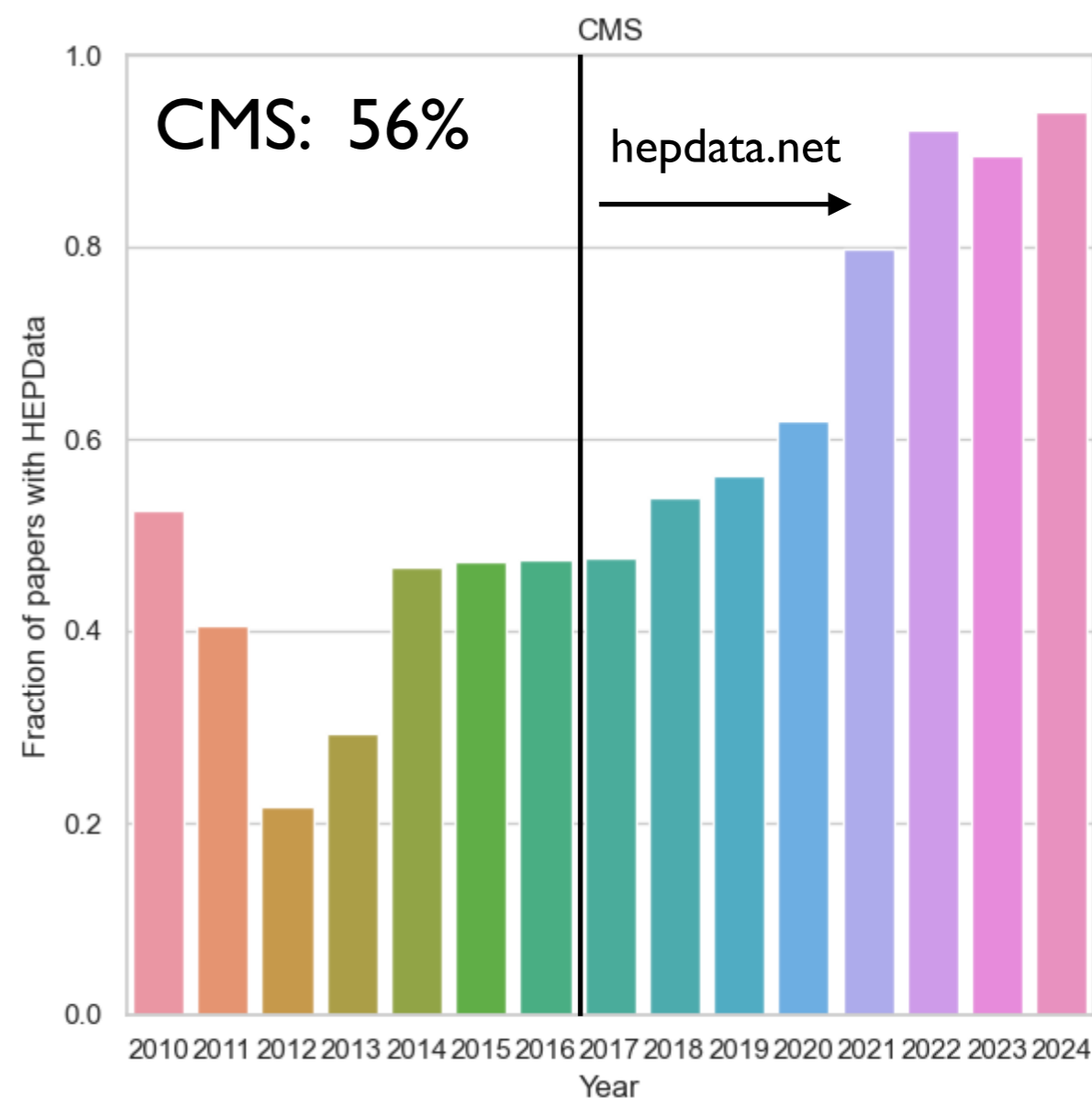
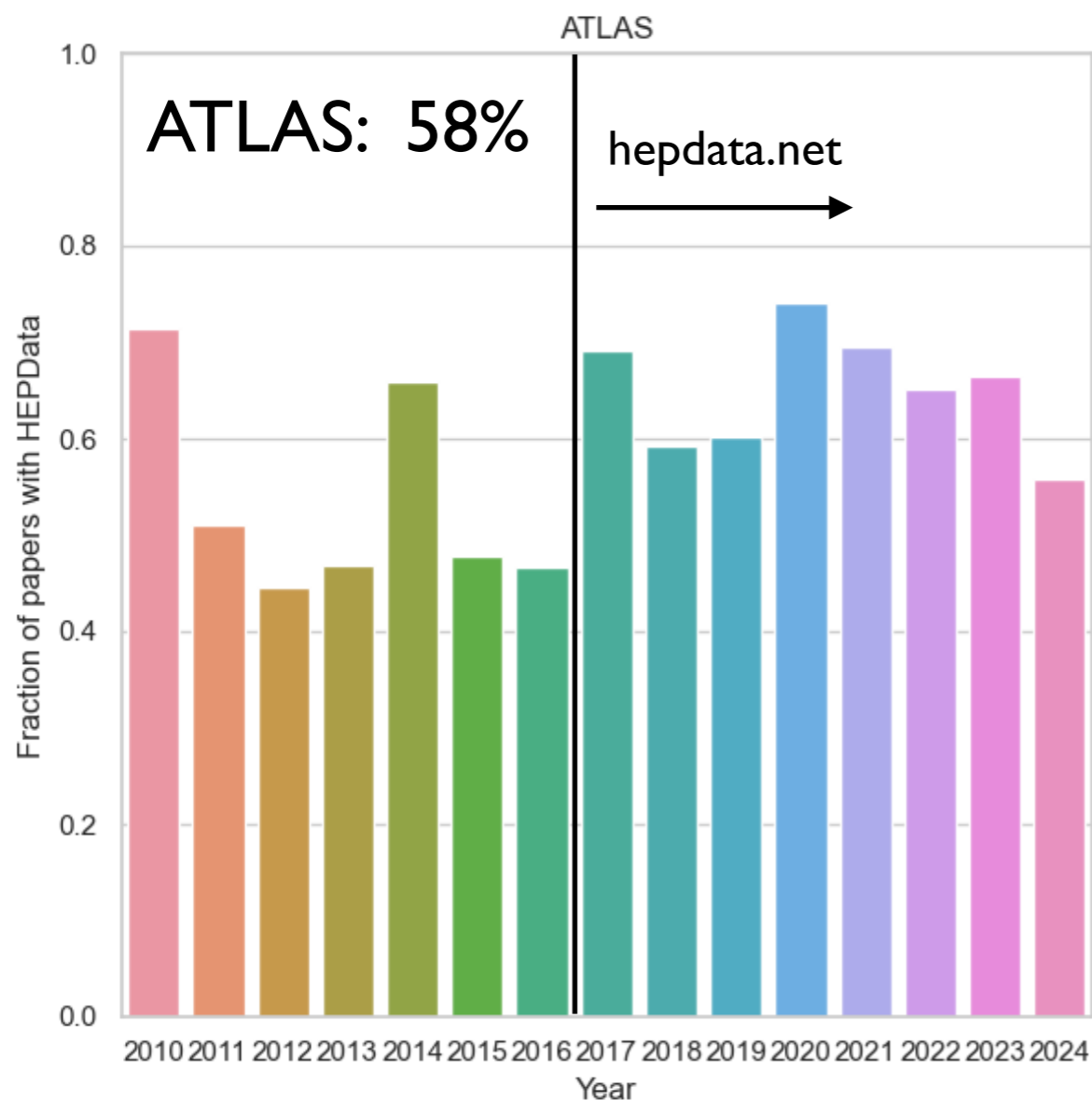
Coverage of LHC publications



- Search INSPIRE for publications with HEPData (GitHub/Binder).

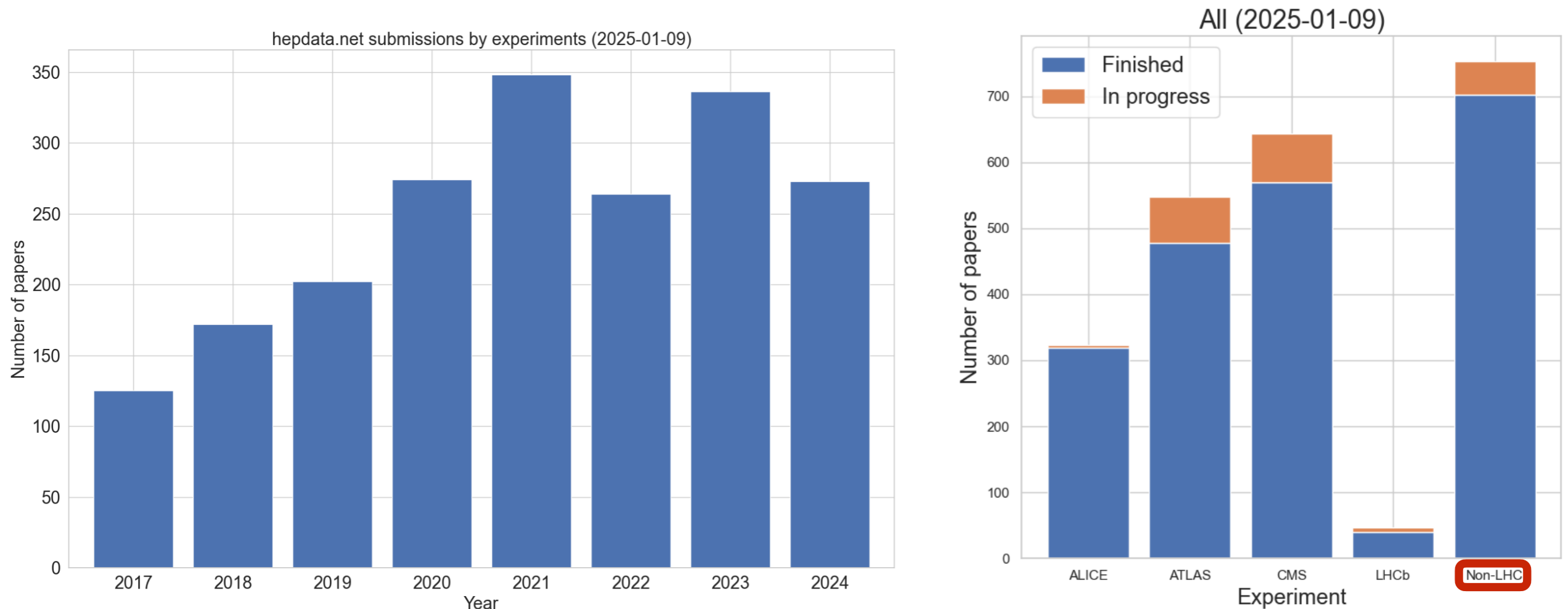
Coverage of ATLAS/CMS publications

LHC publications with HEPData records (2025-01-09)



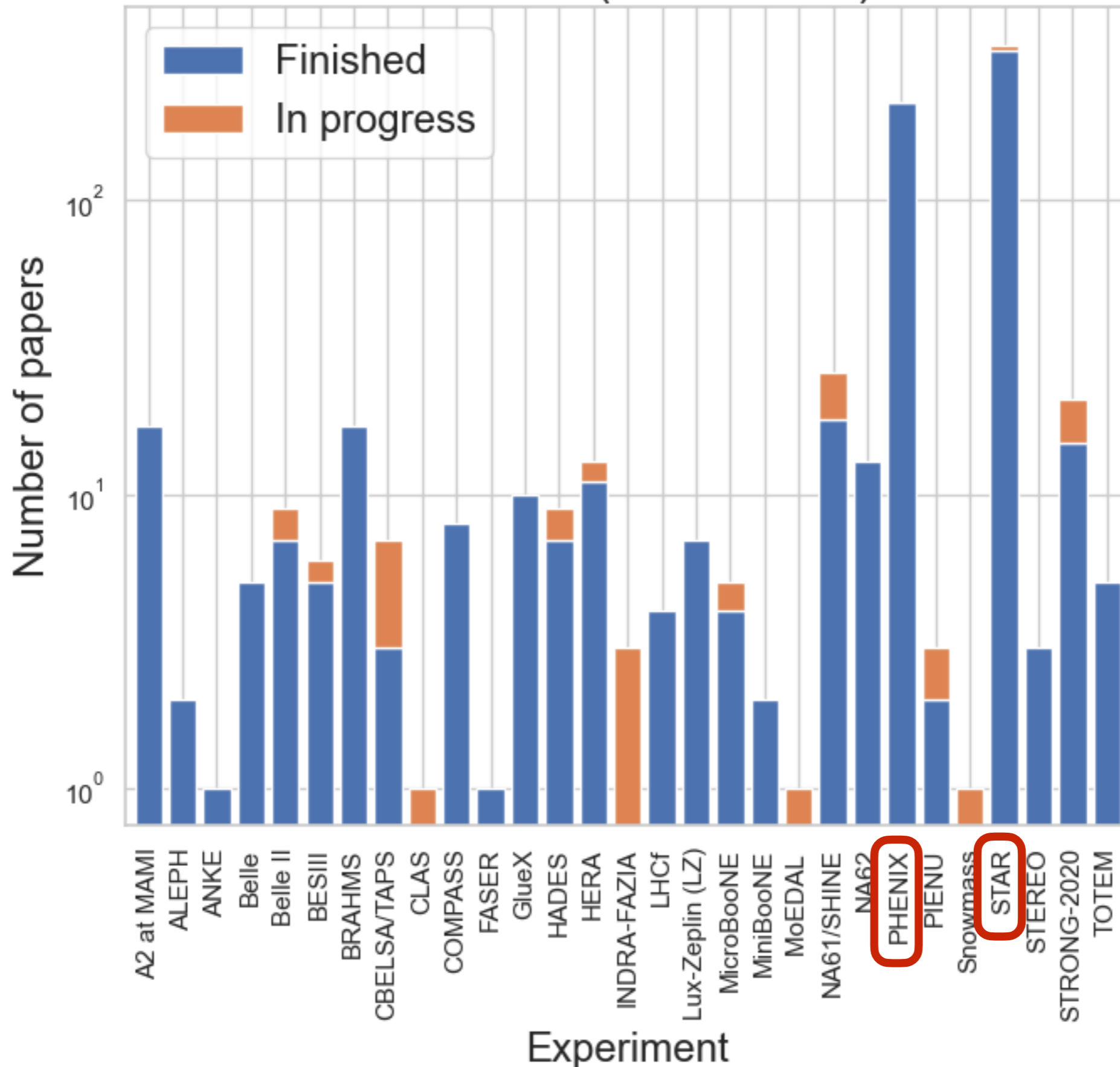
- ALICE (93%) and LHCb (12%) fairly consistent across years.

Submissions via hepdata.net (2017-)



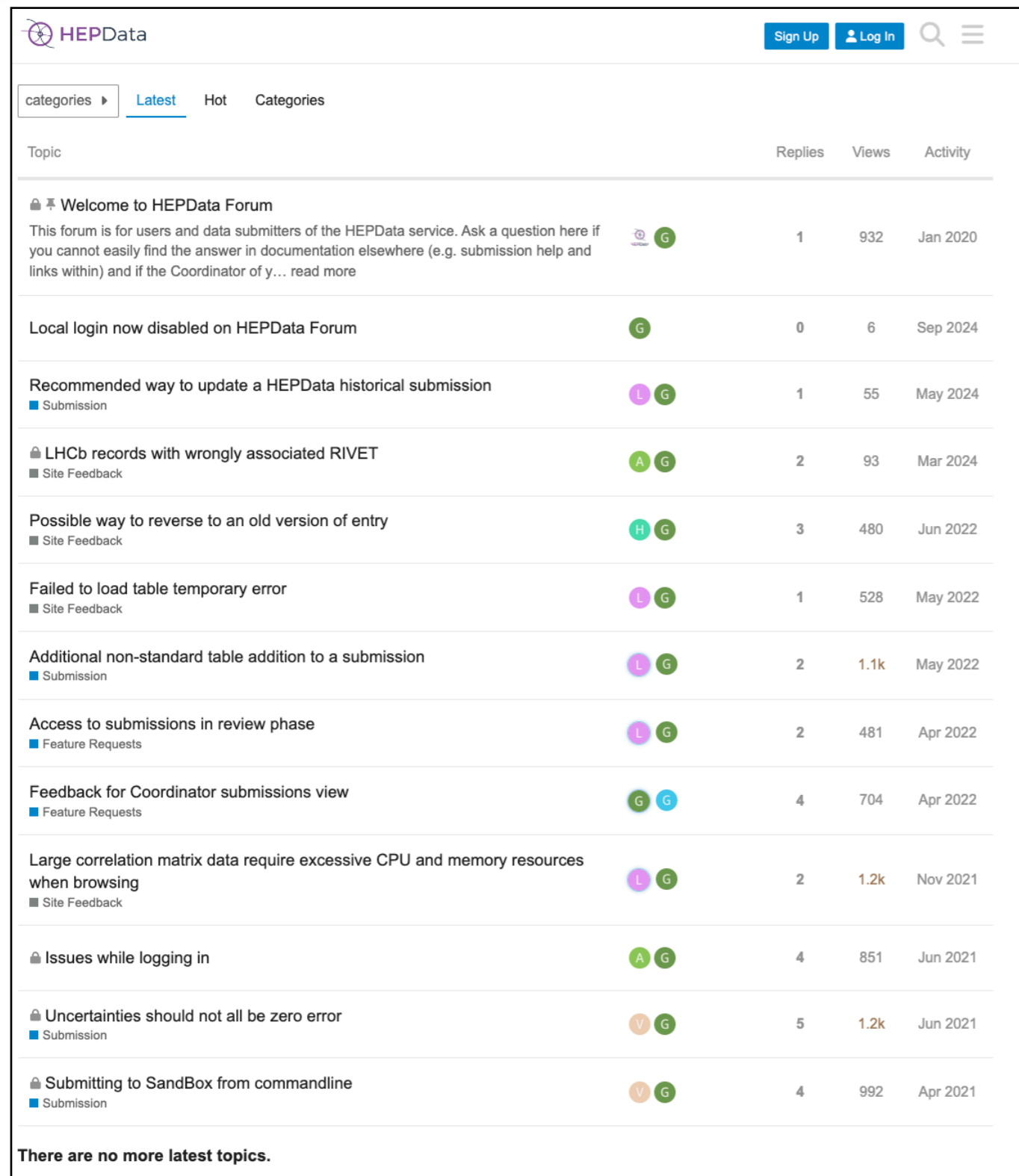
- Increase until 2020, then around 300 submissions/year.
- Large number (702) of completed non-LHC submissions.
- More plots and statistics available in a Jupyter notebook.

Non-LHC (2025-01-09)



- PHENIX ([BNL News](#)) and STAR both with 93% coverage.

hepdata-forum.cern.ch



The screenshot shows the HEPData forum interface. At the top, there are navigation links for 'Sign Up' and 'Log In', along with a search icon and a menu icon. Below this, there are tabs for 'categories', 'Latest', 'Hot', and 'Categories'. The main content area is a table of forum topics. The table has columns for 'Topic', 'Replies', 'Views', and 'Activity'. The topics listed are:

Topic	Replies	Views	Activity
Welcome to HEPData Forum This forum is for users and data submitters of the HEPData service. Ask a question here if you cannot easily find the answer in documentation elsewhere (e.g. submission help and links within) and if the Coordinator of y... read more	1	932	Jan 2020
Local login now disabled on HEPData Forum	0	6	Sep 2024
Recommended way to update a HEPData historical submission Submission	1	55	May 2024
LHCb records with wrongly associated RIVET Site Feedback	2	93	Mar 2024
Possible way to reverse to an old version of entry Site Feedback	3	480	Jun 2022
Failed to load table temporary error Site Feedback	1	528	May 2022
Additional non-standard table addition to a submission Submission	2	1.1k	May 2022
Access to submissions in review phase Feature Requests	2	481	Apr 2022
Feedback for Coordinator submissions view Feature Requests	4	704	Apr 2022
Large correlation matrix data require excessive CPU and memory resources when browsing Site Feedback	2	1.2k	Nov 2021
Issues while logging in	4	851	Jun 2021
Uncertainties should not all be zero error Submission	5	1.2k	Jun 2021
Submitting to SandBox from commandline Submission	4	992	Apr 2021

At the bottom of the table, it says "There are no more latest topics."

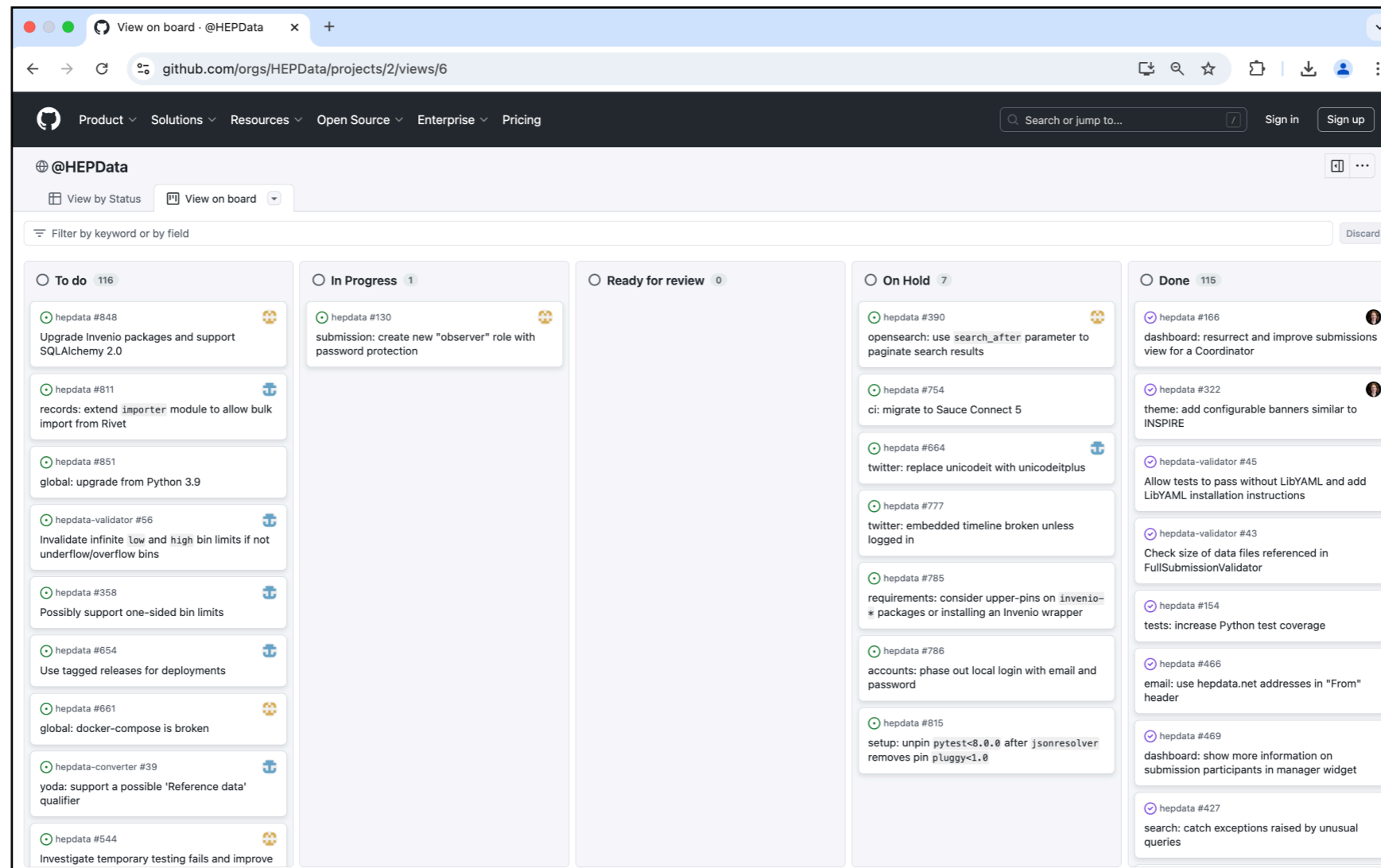
- Request for forum at last HEPData AB meeting.
- Discourse instance set up later in January 2020.
- **Only 13 posts in 5 years.**
- Posts by users mostly for temporary issues that were quickly resolved.
- Users often prefer to email for support than post on a public forum.
- Maybe post to advertise new HEPData features?

Research Software Healthcheck

- Service provided by the Software Sustainability Institute.
- Generated report from online sustainability evaluation.
- Healthcheck application submitted on 12th November 2019.
- Skype with Steve Crouch and James Graham on 10/01/2020.
- Evaluation report received on 25/03/2020, with summary:
*“The project appears in general to be well managed, structured and developed, with only one major concern - **lack of support for Python3.**”*
- Migrated from **Python 2.7** to **Python 3.6** later in 2020, then to **Python 3.9** in 2022. **Python 3.9** reaches EOL in 10/2025.

GitHub project (02/2022-)

<https://github.com/orgs/HEPData/projects>



- HEPData/hepdata: **241** closed issues since 01/2020, but **284** created.
- Travis CI → GitHub Actions (with Sauce Labs) and Dependabot.

Infrastructure updates

- Migration in 2020 from Puppet VMs to Docker/Kubernetes.
- Data files migrated from CERN EOS to CephFS storage.
- Kubernetes configuration specified in private GitHub repo.
- Separate **QA** environment for testing before *production*.
- Currently PostgreSQL v14.10 and OpenSearch v2.15.0.
- Attempts to mitigate problems of large submissions, e.g.:
 - Increase web server request timeout from **1** to **5** minutes.
 - Separate “Upload and Process” so *Process* asynchronous.
 - Impose limit of **50 MB** for uploaded archive file.
 - Impose limit of **10 MB** for *any* YAML data file in archive.

NEW

from Dec
2023

Deferred loading of large files

- Validator restricts YAML data files to be smaller than **10 MB**.
- YAML data files **1-10 MB** can be slow to load in browser.
- Introduce deferred table loading if YAML data file **> 1 MB**:
- On resource file landing page if text file **> 1 MB**, only display “Download” button without trying to display in browser:

This table is too large to load automatically.
The table size is 9.25 MB, which is greater than our threshold of 1.00 MB.

Load Table

workspace_tHu.json [10.17182/hepdata.150998.v1/r1](https://doi.org/10.17182/hepdata.150998.v1/r1)

License: [CC0](#)

Full likelihood of the tHu fit in the HistFactory JSON format described in ATL-PHYS-PUB-2019-029

This file (2.21 MB) is larger than our loading threshold (1.00 MB), and is only available for download below.

Download via DOI: `curl -OJLH "Accept: application/json" https://doi.org/10.17182/hepdata.150998.v1/r1`

Download

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>







New INSPIRE-HEP (March 2020)

- More work by Giuseppe De Laurentis in summer 2020:
 - Get publication information in new JSON format using INSPIRE REST API instead of old MARCXML format.
 - Automatic updates of publication information displayed on HEPData, e.g. if journal reference added to INSPIRE record.
- Provided an endpoint hepdata.net/search/ids.
- Search INSPIRE for records with HEPData using query: external system identifiers.schema:HEPData
- “**datasets**” links added to INSPIRE from January 2021:

First Measurement of $a_2^0(1320)$ Polarized Photoproduction Cross Section #1

GlueX Collaboration • F. Afzal et al. (Jan 6, 2025)

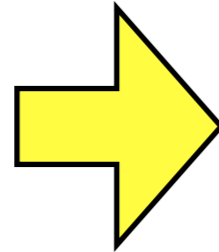
e-Print: [2501.03091](#) [nucl-ex]

 pdf  cite  **datasets**  claim  reference search  0 citations

Data output formats

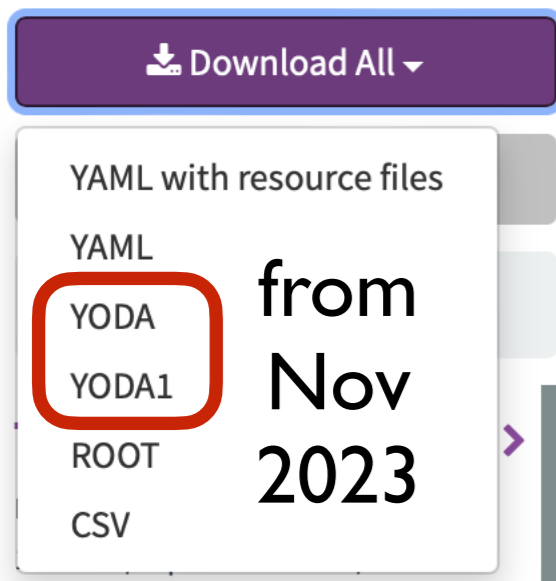
hepdata.net/formats

YAML: native
HEPData format.



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

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



- **NEW** “YODA” now gives new YODA2 format, for use with Rivet version 4 (released Feb 2024).
- Legacy YODA format still available as YODA1.
- Thanks to [Chris Gütschow \(UCL\)](#) for work on implementing the YAML → YODA2 conversion.

Links to analysis code

- JSON file maps INSPIRE IDs to Rivet analysis names:

```
{ "100016" : [ "GAMMAGAMMA_1975_I100016" ], ...,  
  "954993" : [ "ATLAS_2011_I954993" ] }
```

- Nightly task parses JSON file and adds new analyses.

- Search query analysis:rivet (from May 2022).

- Extended to code from other analysis frameworks:

analysis:MadAnalysis (from Oct 2023)

analysis:SModels (from Nov 2024)

analysis:Combine (from Dec 2024)

- Allow bulk subscription to record update notifications.

Additional resource files

The screenshot shows the HEPData website interface. A modal window titled "Additional Publication Resources" is open, displaying a list of resources. The resources are organized into a grid. Each resource card includes a description, license (CC0), and a DOI. The DOIs are highlighted with red circles. The resources are:

- Param card (SLHA file) for Gbb 2000, 1000 model, location: param_card_376017.dat. License: CC0. DOI: 10.17182/hepdata.95928.v2/r2. Buttons: Landing Page, Download.
- Param card (SLHA file) for Gtb 2200, 600 model, location: param_card_376093.dat. License: CC0. DOI: 10.17182/hepdata.95928.v2/r3. Buttons: Landing Page, Download.
- C++ File. Description: Code for NN and CC regions in SimpleAnalysis, location: ANA-SUSY-2018-30.cxx. License: CC0. DOI: 10.17182/hepdata.95928.v2/r4. Buttons: Landing Page, Download.
- HistFactory File. Description: Archive of full likelihoods in the HistFactory JSON format. License: CC0. DOI: 10.17182/hepdata.95928.v2/r5. Buttons: Landing Page, Download.

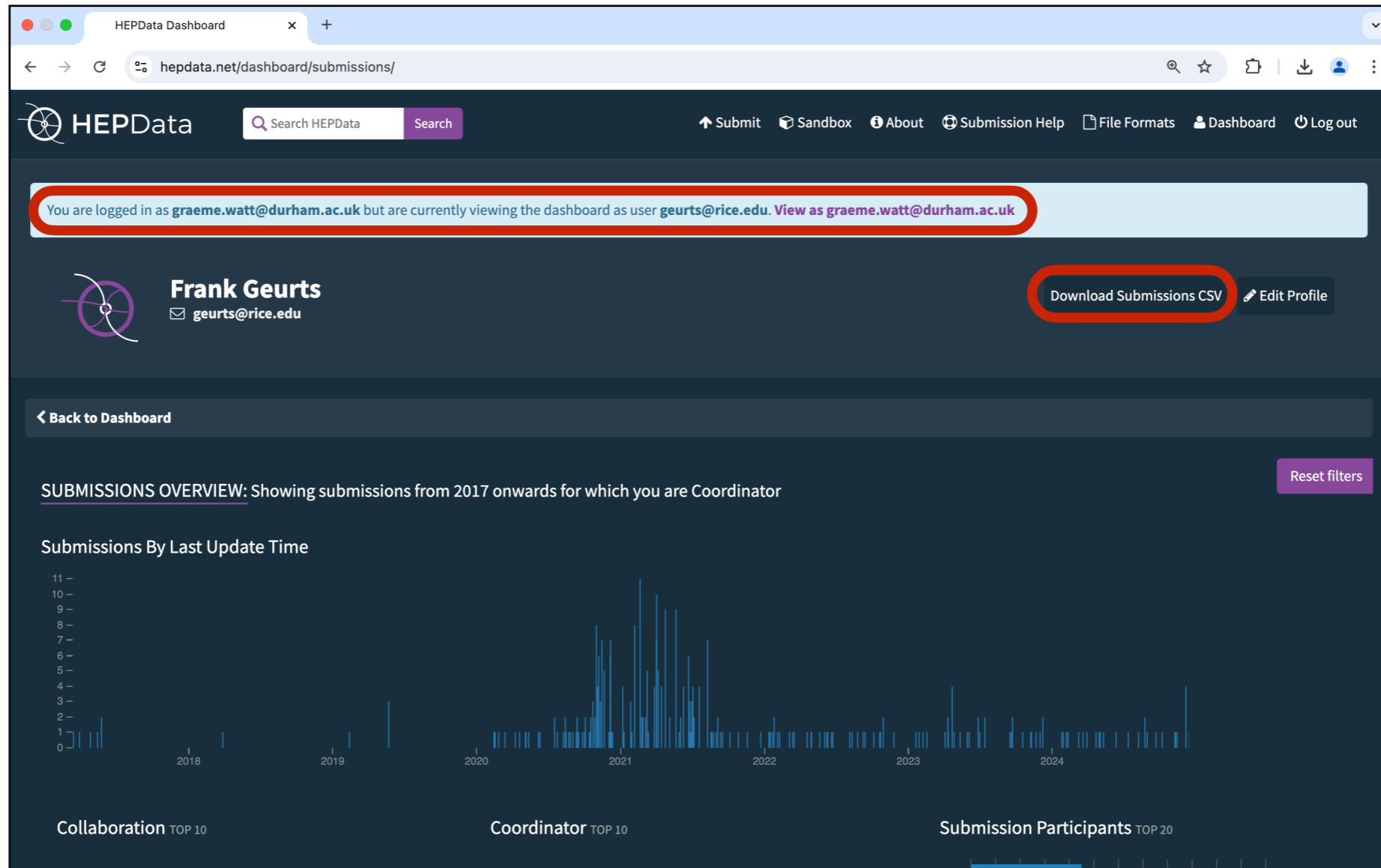
The HistFactory File resource is highlighted with a red box. The background shows the HEPData search results for "analysis:HistFactory".

- DOIs now minted via DataCite for additional resource files.
- Search query analysis:HistFactory (from May 2022).
- Search query analysis:NUISANCE (from July 2024).

NEW

from April
2022

Submissions Overview



- Interactive graphical overview of a **Coordinator's** submissions.
- CSV export for further offline analysis (e.g. in a spreadsheet).
- Ability for **Admin** user to view as a particular **Coordinator**.

NEW

from Aug
2023

Bidirectional linking

- Suggestion by Jon Butterworth in December 2022. Technical implementation by Jordan Byers in 2023.
- Enable bidirectional links *between* HEPData **tables** possibly in different records in `submission.yaml`:

```
related_to_table_dois:
```

```
- 10.17182/hepdata.12345.v1/t2  
- 10.17182/hepdata.67890.v3/t4
```

or use hepdata_lib

- Similar bidirectional links *between* HEPData **records**:

```
related_to_hepdata_records:
```

```
- 12345  
- 67890
```

or use hepdata_lib

NEW

from Aug
2023

Bidirectional linking

HEPData | ATLAS | 2024 | Me

hepdata.net/record/ins2771257

Search HEPData

Table 12 10.17182/hepdata.151815.v1/t12

License: CC0

This table is related to:

- Table 4

This table is referred to by:

- Table 18
- Table 24

Measurements of the production cross-section for a Z boson in association with b- or c-jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS collaboration

Aad, Georges, Aakvaag, Erlend, Abbott, Braden Keim, Abdelhameed, Sara, Abeling, Kira, Abicht, Nils Julius, Abidi, Halder, Aboelela, Mohammed, Aboulhorma, Asmaa, Abramowicz, Halina

Eur.Phys.J.C 84 (2024) 984, 2024.

https://doi.org/10.17182/hepdata.151815

Journal INSPIRE Resources

Abstract (data abstract)

This paper presents a measurement of the production cross-section of a Z boson in association with b- and c-jets, in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment at the Large Hadron Collider using data corresponding to an integrated luminosity

Table 15

10.17182/hepdata.151815.v1/t15

Correction factors from parton level done with flavour dressing algorithms as seen in Ref. [2] of the paper to hadron...

Table 16

10.17182/hepdata.151815.v1/t16

Correction factors from parton level done with flavour dressing algorithms as seen in Ref. [2] of the paper to hadron...

Table 17

10.17182/hepdata.151815.v1/t17

Correction factors from parton level done with flavour dressing algorithms as seen in Ref. [2] of the paper to hadron...

cmenergies

13000.0

reactions

P P -> Z0 HEAVY-FLAVOR

Visualize

RE P P -> Z0 HEAVY-FLAVOR

SQRT(S) 13000 GEV

Distribution Correction Factors

HEPData | HADES | 2023 | Pr

hepdata.net/record/ins152804

Search HEPData

Table 12 10.17182/hepdata.152804.v1/t12

License: CC0

This record is related to:

- 102468

Proton, deuteron and triton flow measurements in Au+Au collisions at $\sqrt{s_{NN}} = 2.4$ GeV

The HADES collaboration

Adamczewski-Musch, J., Arnold, O., Behne, C., Belounnas, A., Berger-Chen, J.C., Blanco, A., Blume, C., Böhmer, M., Bordalo, P., Chlad, L.

Eur.Phys.J.A 59 (2023) 80, 2023.

https://doi.org/10.17182/hepdata.152804

Journal INSPIRE Resources

Abstract

High precision measurements of flow coefficients v_n ($n = 1 - 4$) for protons, deuterons and tritons relative to the first-order spectator plane have been performed in Au+Au collisions at $\sqrt{s_{NN}} = 2.4$ GeV with the High-Acceptance Di-Electron Spectrometer (HADES) at the SIS18/GSI. Flow coefficients are studied as a function of transverse momentum p_T and rapidity y_{cm} over a large region of phase space and for several classes of collision centrality. A clear mass hierarchy is found for the slope of v_1 , $dv_1/dy|_{y=0}$ where y' is the scaled

Figure 5 Event Plane Resolution 10pct

10.17182/hepdata.152804.v1/t11

Data from Figure 5 (square marker), located on page 6.

The resolution \mathcal{R}_n of the first-order spectator event plane for flow coefficients of different orders n as a function of the event centrality (Adamczewski-Musch:2020iio). The circles correspond to centrality intervals of 5% width and the squares to 10% width (curves are meant to guide the eye).

Figure 5 Event Plane Resolution 5pct

Data from Figure 5 (circle marker), located on page 6.

10.17182/hepdata.152804.v1/t12

The resolution \mathcal{R}_n of the first-order spectator event plane for flow coefficients of different orders n as a function of...

Figure 8 v1 ycm Protons

Data from Figure 8 (left column top row), located on page 10.

10.17182/hepdata.152804.v1/t13

The flow coefficients v_1 , v_2 , and

observables

EP

phrases

Flow Event Plane Resolution

Centrality [%]	R_1	R_2	R_3	R_4	R_5	R_6
5.0	0.64812	0.2978	0.11081	0.035458	0.010092	0.0
15.0	0.84618	0.57111	0.32662	0.1642	0.07434	0.0
25.0	0.88566	0.65032	0.41238	0.23228	0.11857	0.0
35.0	0.87041	0.61814	0.37602	0.20226	0.098316	0.0

Visualize

related_to_table_dois:

- 10.17182/hepdata.151815.v1/t4

related_to_hepdata_records:

- 102468

- Four ATLAS records linking tables: 141650, 138878, 149990, 151815.
- One HADES record 152804 (2024) links to 102468 (2021).
- Links are automatically *bidirectional*: only specify one direction.

NEW

from May 2024

Searching for resources

Searching resources by field

Text-based description searching:

`resources:"Created with hepdata_lib"`

Quotes force a full match.

Resource-type searching:

`resources.type:png`

Examples: png, html, github, zenodo etc.

Searching for specific URLs:

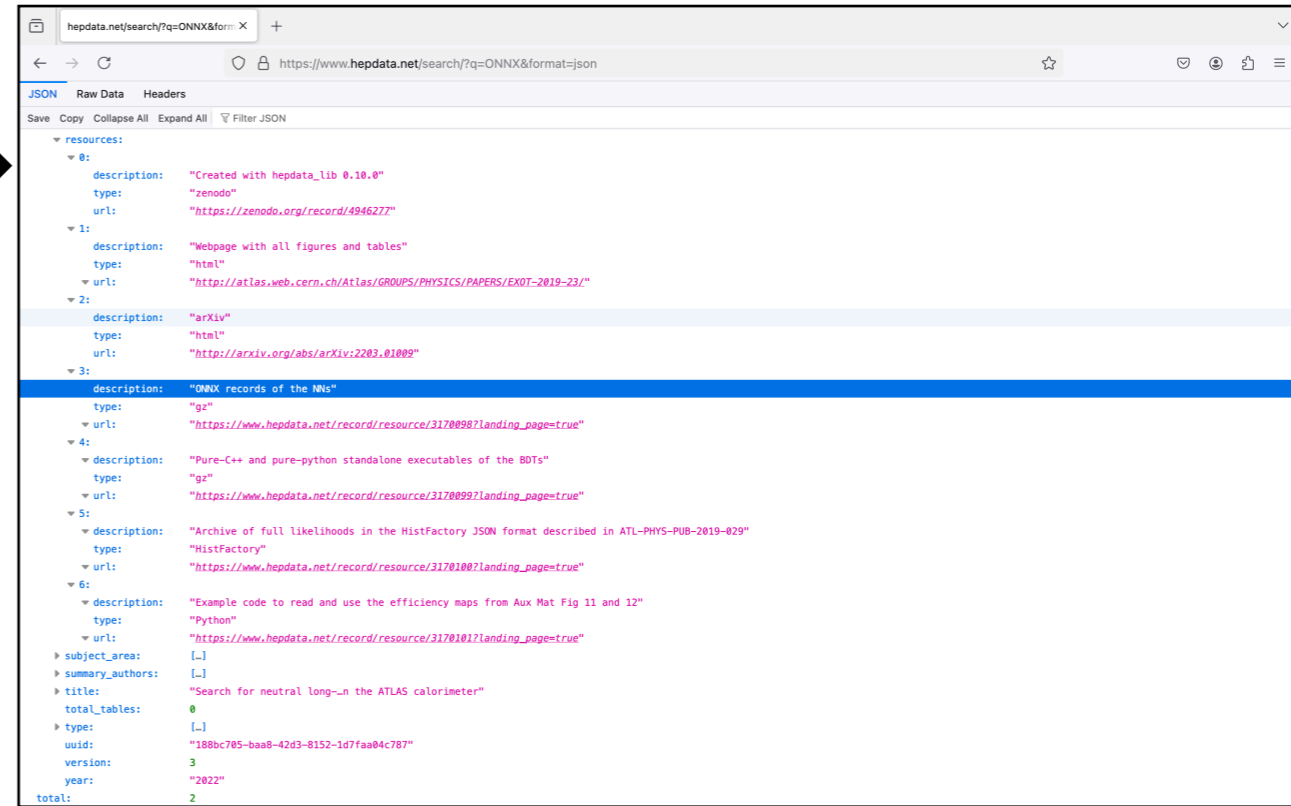
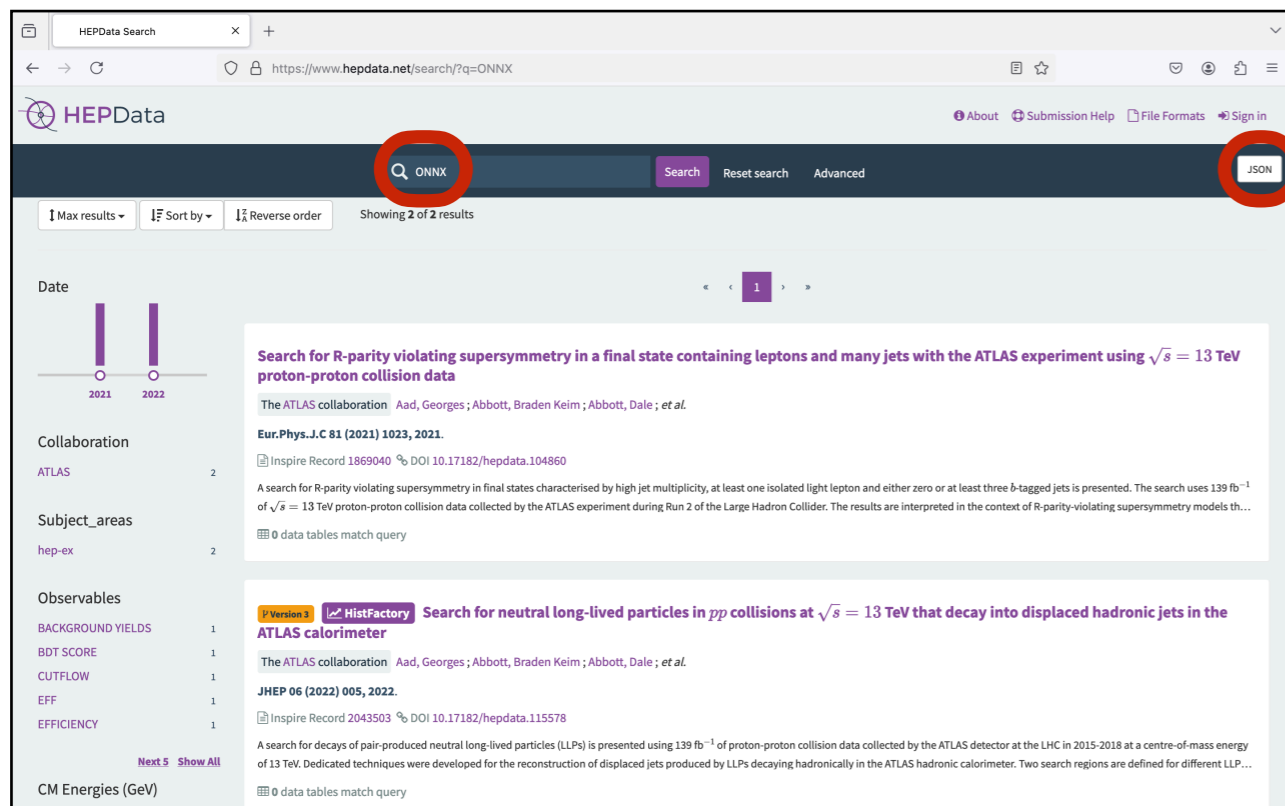
`resources.url:atlas.web.cern.ch`

- Additional resource metadata now indexed for searching.

NEW

from May 2024

Resource information in JSON



```
import requests
query = 'ONNX'
url = f'https://www.hepdata.net/search/?q={query}&format=json'
request = requests.get(url).json()
for result in request['results']:
    for resource in result['resources']:
        if query in resource['description']:
            print(result['inspire_id'])
            print(resource['description'])
            print(resource['type'])
            print(resource['url'].replace('landing_page', 'view'))
            print()
```

1869040
ONNX files for the neural networks for the EWK analysis
tgz
<https://www.hepdata.net/record/resource/2677521?view=true>

2043503
ONNX records of the NNs
gz
<https://www.hepdata.net/record/resource/3170098?view=true>

● **Example:** get download links of ONNX files from Python.

Summary

Email: info@hepdata.net

Forum: hepdata-forum.cern.ch

- **HEPData** is *the* repository for publication-related HEP data.
- *Caveats:* design restricts size (\leq MB) and format (mostly tabular).
- Widely used by HEP community: **~4 million** page views per year.
- Infrastructure via CERN, development and support via Durham.
- Open development process via <https://github.com/HEPData> .
- Backlog of (101) issues: <https://github.com/HEPData/hepdata/issues>
- **(Near) future plans:**
 - Create new “observer” role with password protection ([#130](#)).
 - Bulk import 780 submissions from [Rivet](#) to HEPData ([#811](#)).
 - Migrate to [SQLAlchemy](#) 2.0 to use latest [Invenio](#) ([#848](#)).