HEPData in STAR: current status & proposal

Frank Geurts

RICE UNIVERSITY

Outline

- Introducing HEPData
- Accessing/Uploading HEPData records

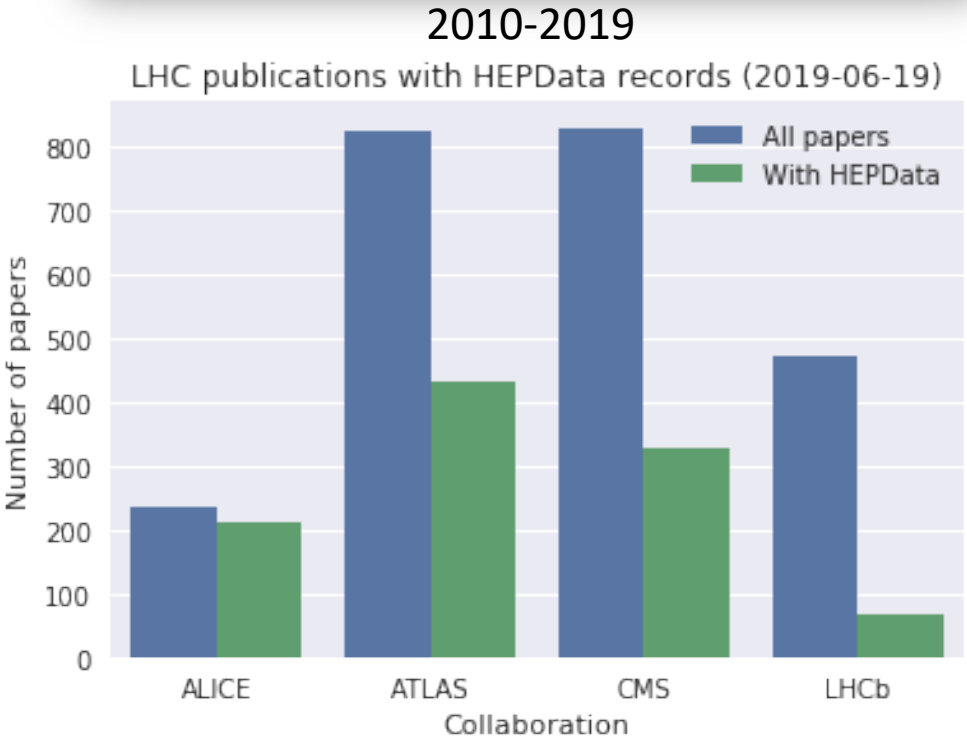
  where are we now?
- STAR & HEPData
- HEPData submission protocol

  … and what is next?
- Involve GPCs and opportunities for service work

What is HEPData?

- *Repository for publication-related High-Energy Physics data*
  - previously known as the Durham Database

➢Built up over the past four decades as a unique open-access repository
  - Has all published LHC particle *and* heavy-ion data



INSPIRE search query:
- hep-ex or nucl-ex
- Published in a journal
- Not conference paper

ALICE: 90%
ATLAS: 52%
CMS: 39%
LHCb: 14%

G.Watt, 5th workshop on (re-)interpreting the results of new physics at the LHC, Apr.'19

More than 4 decades …

Currently (March 14, 2020): 8999 publications with 84077 tables

Earliest entry: Durbin, Loar, and Steinberger, Phys. Rev. 84 (**1951**)

PHYSICAL REVIEW VOLUME 84, NUMBER 3 NOVEMBER 1, 1951

Letters to the Editor

PUBLICATION of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is five weeks prior to the date of issue. No proof will be sent to the authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not exceed 600 words in length.

The Absorption of Pions by Deuterons*

R. DURBIN, H. LOAR, AND J. STEINBERGER
Columbia University, New York, New York
(Received September 17, 1951)

THE study of the absorption of positive pions by deuterons, which, together with its inverse, has been used to determine the spin of the meson,[1,2] has been extended. We present here results on the energy dependence of the angular distribution and of the total cross section, in the hope that they will prove useful in the attempt to understand the pion-nucleon interaction.

The reaction is $\pi^+ + d \rightarrow p + p$. The reaction rate is determined by the rate of recoil proton pairs in coincidence with incoming mesons, incident successively on water and heavy water targets. The experimental technique has already been described.[1] The water cells are 1″ thick along the beam direction. The energy dispersion due to target thickness and meson beam inhomogeneity is approximately ±7 Mev. The angular dispersion is approximately ±14°. Despite this poor resolution the coincidence rate of recoil protons is only ~1/minute in the Nevis meson beam of ~500/sec. The uncertainties in the result are almost entirely statistical; the geometrical factors have been calculated with greater accuracy, and the combined uncertainties in counting efficiency and beam composition are less than 10 percent.

The results are given in Table I. All parameters as well as differential cross sections are in the center of mass system. The recoil angles listed are averaged over the angular dispersion. This affects only the angles near 90°, since, because of the symmetry, 90° is the extreme angle, and larger angles are recorded as smaller angles. When the proton counter axes are set at 90° relative to the meson beam, the average detection angle is 83°.

It is possible to represent the angular dependence as $\alpha + \beta \cos^2\theta$, since meson angular momenta with respect to the deuteron of more than two Planck units should not contribute appreciably at these energies. The best fits to such a distribution are given in Table II, together with absolute cross sections.

In Table II we have also included the results of Cartwright, Richman, Wilcox, and Whitehead.[3] The angular distribution found at Berkeley and our results are only in fair agreement.

There is no large change in the angular distribution in the energy range 25–53 Mev for the incident meson in the center of mass system. The total cross section increases by a factor 2.25±0.32.

To see the implication of this result on the meson nucleon interaction, it is necessary to separate the effects of the nuclear binding in initial and final states as well as the kinematical factors. The effects of the binding on the angular distribution are complicated, and are discussed in the following note.[4] The kinematical factors are

TABLE I. Differential cross sections for the absorption of pions by deuterons in the center of mass system. Rms statistical errors are given.

E_m = 25 Mev		E_m = 40 Mev		E_m = 53 Mev	
θ_{cm}	$d\sigma/d\Omega$ cm²/sterad	θ_{cm}	$d\sigma/d\Omega$	θ_{cm}	$d\sigma/d\Omega$
29°	8.5±1.1 ×10⁻²⁸	30°	17.4±2.6 ×10⁻²⁸	30°	20.1±4 ×10⁻²⁸
35°	7.3±1.6	50°	9.2±1.5	60°	5.7±1.8
45°	7.5±2.2	70°	8.4±2.4	83°	5.5±2.1
50°	7.1±0.9	83°	4.0±1.7	···	···
82°	1.7±0.6	···	···	···	···

TABLE II. Total cross sections and best fit angular distributions for the data in Table I.

E_{cm}	Angular distribution	Total cross section
21.5 Mev*	10.7 $(\cos^2\theta + 0.07) \times 10^{-28}$ sterad/cm²	2.8±0.8 ×10⁻²⁷ cm²
25 Mev	9 $(\cos^2\theta + 0.22) \times 10^{-28}$ sterad/cm²	3.1±0.3 ×10⁻²⁷ cm²
40 Mev	18 $(\cos^2\theta + 0.2) \times 10^{-28}$ sterad/cm²	6.1±0.6 ×10⁻²⁷ cm²
53 Mev	21.5 $(\cos^2\theta + 0.18) \times 10^{-28}$ sterad/cm²	7.0±0.7 ×10⁻²⁷ cm²

* We include the result of the Berkeley group on the inverse reaction, inverted on the basis of the law of detailed balance and zero spin for the pion.

K = momentum space/(relative velocity of incoming nucleons ×total meson energy).

The factor 1/total meson energy is not strictly a kinematical factor; it is due to the normalization of the meson wave and is included because it is omitted in the theoretical analysis which follows this letter.

$$k_{53}/k_{25} = 0.71.$$

The average square of the matrix element for the process therefore increases by the factor 2.3/0.71 = 3.25±45 in the energy range 25–53 Mev. This is a considerable increase, especially since the effects of binding also decrease the cross section at higher relative to lower meson energy. This effect is approximately

$$|f(k_{53})/f(k_{25})|^2,$$

where $f(k_{53})$ is the fourier amplitude of the deuteron or the diproton wave function for the momentum of one of the recoil protons resulting in the absorption of a 53-Mev meson. These fourier amplitudes decrease at least as $1/k^2$ and probably more nearly as $1/k^3$ for such large momenta. If this factor is also taken into account, then the meson-nucleon interaction must increase by a factor $\sqrt{4.5} - \sqrt{6}$ in the meson energy interval 25–53 Mev. The meson momentum increases by the factor 1.53 in this interval so that the meson nucleon interaction must increase approximately as the square of the momentum of the meson. This is of course a stronger momentum dependence than the linear dependence predicted in pseudoscalar theory with pseudovector coupling. Other theories predict an even weaker dependence and are also in conflict with other experiments. The steep energy dependence is probably of the same origin as the steep excitation function in neutral photomeson production[5,6] and may be due to some resonance effect, as has already been suggested in connection with the photomeson experiments.[1,5]

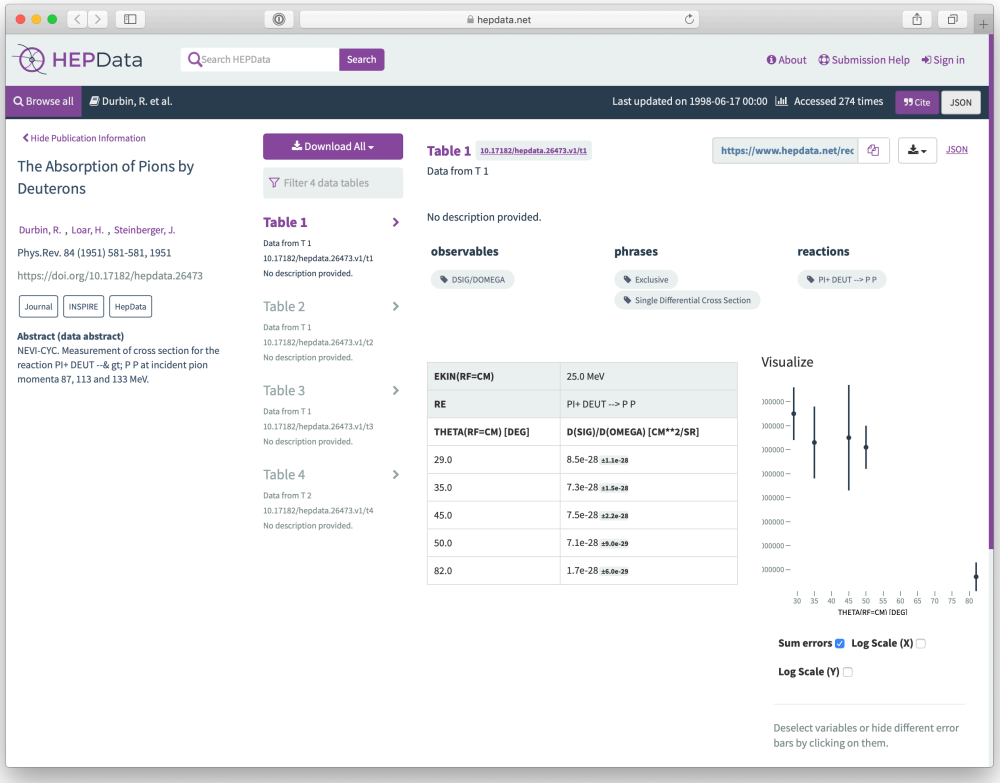
We wish to thank the operating crew of the Nevis cyclotron, under the direction of J. Spiro, for the bombardments.

* Research sponsored by a joint program of the ONR and AEC.
[1] Durbin, Loar, and Steinberger, Phys. Rev. **83**, 646 (1951).
[2] Clark, Roberts, and Wilson, Phys. Rev. **83**, 649 (1951).
[3] M. Whitehead and C. Richman, Phys. Rev. **83**, 855 (1951).
[4] Chew, Goldberger, Steinberger, and Yang, Phys. Rev. **84**, 581 (1951).
[5] A. Silverman and M. Stearns, Phys. Rev. **83**, 853 (1951).
[6] Panofsky, Steinberger, and Steller (to be published).
[7] K. Brueckner and K. Case, Phys. Rev. **83**, 1141 (1951).
[8] Y. Fujimoto and H. Miyazawa (to be published).

A Theoretical Analysis of the Process $\pi^+ + d \rightleftarrows p + p$

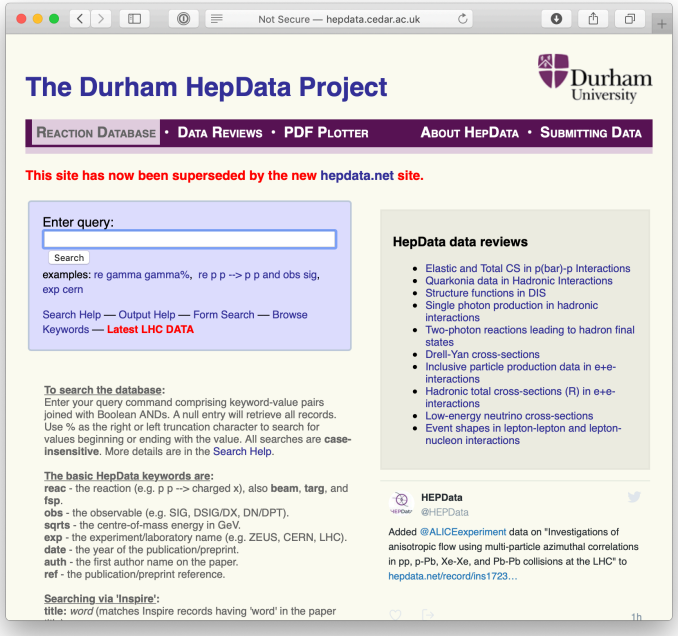
GEOFFREY F. CHEW, *Brookhaven National Laboratory, Upton, New York*
M. L. GOLDBERGER, *Institute for Nuclear Studies, University of Chicago, Chicago, Illinois*
J. M. STEINBERGER, *Columbia University, New York City, New York*
AND
C. N. YANG, *Institute for Advanced Study, Princeton, New Jersey*
(Received September 17, 1951)

THE experiment described in the preceeding letter[1] as well as those on the inverse reaction, the production of mesons in the collision of two protons, may be analyzed in the spirit of

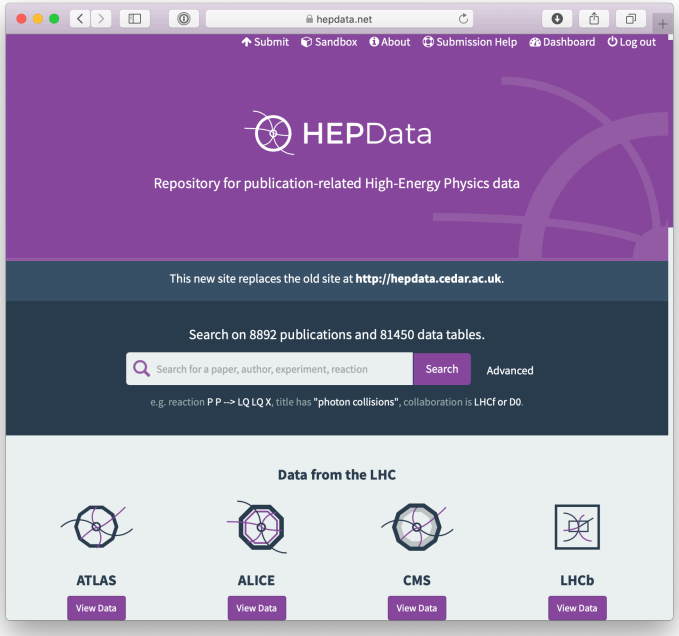


… and its own history

Pre 2016

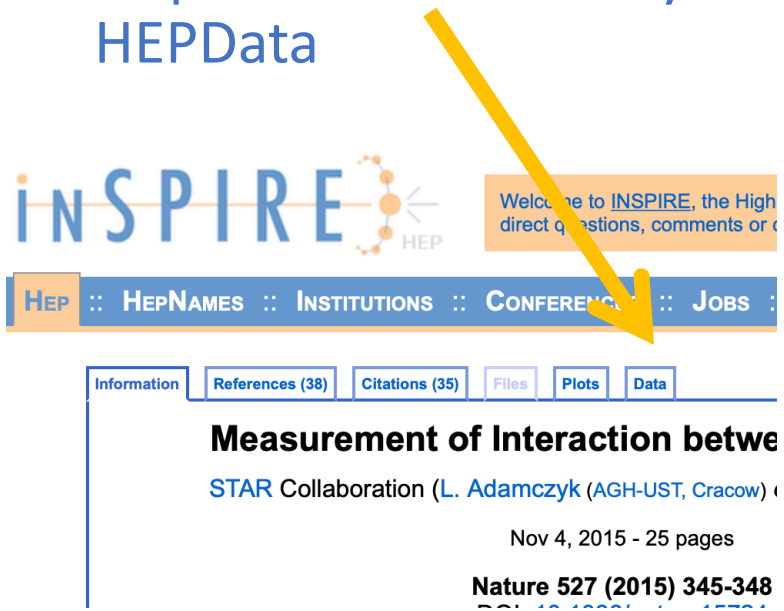


2016+



2016: HEPData transition from Durham to CERN

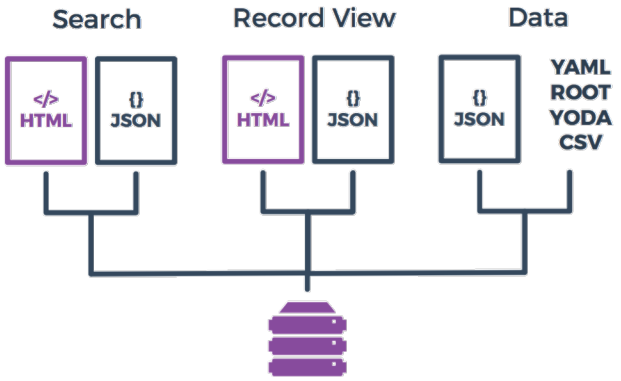
- http://hepdata.net
- based on Invenio, developed in collaboration with INSPIRE and hosted at CERN
- Inspire records directly linked to HEPData



But also …

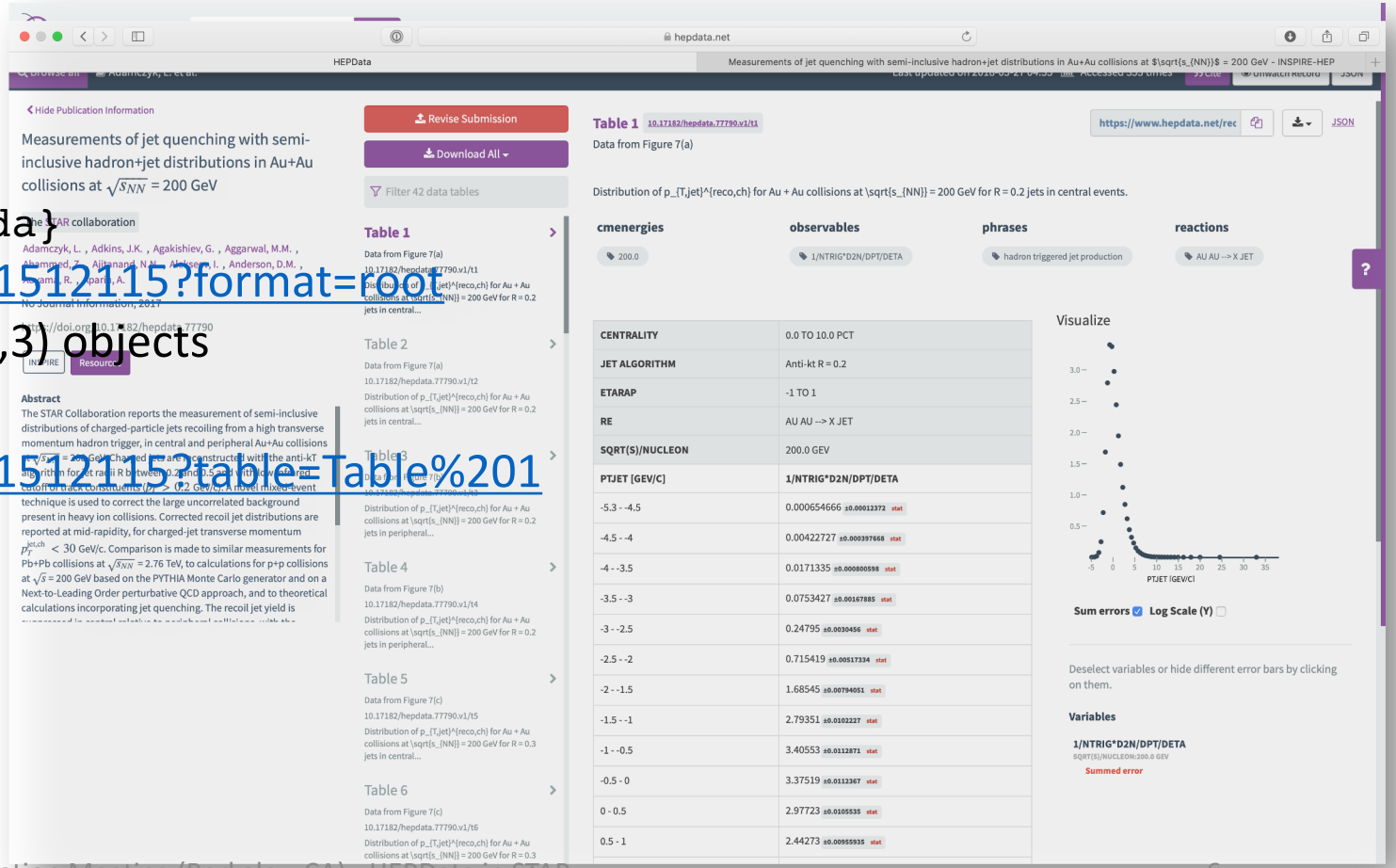
- ➢ Change in upload format
- ➢ Change in submission protocol

Access to HEPData records



Maguire, arXiv:1704.05473

- Use HTML web pages
  - example: 2017 jet quenching paper, PRC96 (2017) 024905
  - https://www.hepdata.net/record/ins1512115
- Programmatic access
  - various output formats
    - `format={csv,json,root,yaml,yoda}`
    - https://www.hepdata.net/record/ins1512115?format=root
    - TGraphAsymmErrors and THxF (x=1,2,3) objects
  - specify what table
    - https://www.hepdata.net/record/ins1512115?table=Table%201
- DOI record for easy citation
  - 10.17182/hepdata.77790.v1/t1



Upload Format

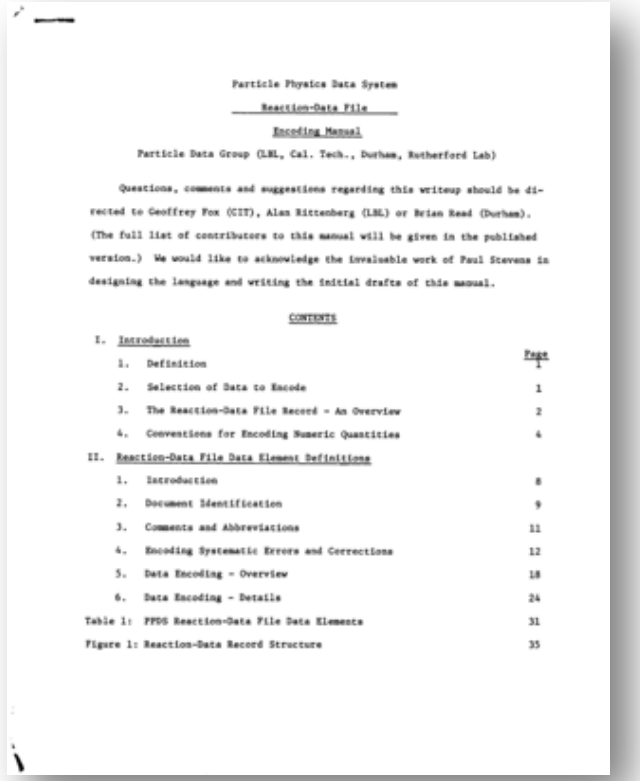
- Old HepData format is based on the mid-70s *reaction-data* files used in BDMS
- New HEPdata based on YAML
  - ~~Yet Another~~ YAML Ain't Markup Language

- Typical submission format
  - One zip file with …
    - submission.yaml
    - Table1.yaml

    …
    - TableN.yaml

More information: https://hepdata-submission.readthedocs.io/

Official *Encoding Manual* (LBL, CalTech,Durham, Rutherford Lab)



Particle Physics Data System

Reaction-Data File

Encoding Manual

Particle Data Group (LBL, Cal. Tech., Durham, Rutherford Lab)

Questions, comments and suggestions regarding this writeup should be directed to Geoffrey Fox (CIT), Alan Rittenberg (LBL) or Brian Read (Durham). (The full list of contributors to this manual will be given in the published version.) We would like to acknowledge the invaluable work of Paul Stevens in designing the language and writing the initial drafts of this manual.

CONTENTS

I. Introduction — Page 1
1. Definition 1
2. Selection of Data to Encode 1
3. The Reaction-Data File Record - An Overview 2
4. Conventions for Encoding Numeric Quantities 4
II. Reaction-Data File Data Element Definitions
1. Introduction 8
2. Document Identification 9
3. Comments and Abbreviations 11
4. Encoding Systematic Errors and Corrections 12
5. Data Encoding - Overview 18
6. Data Encoding - Details 24
Table 1: PPDS Reaction-Data File Data Elements 31
Figure 1: Reaction-Data Record Structure 35

YAML … source: Wikipedia

History and name [edit]

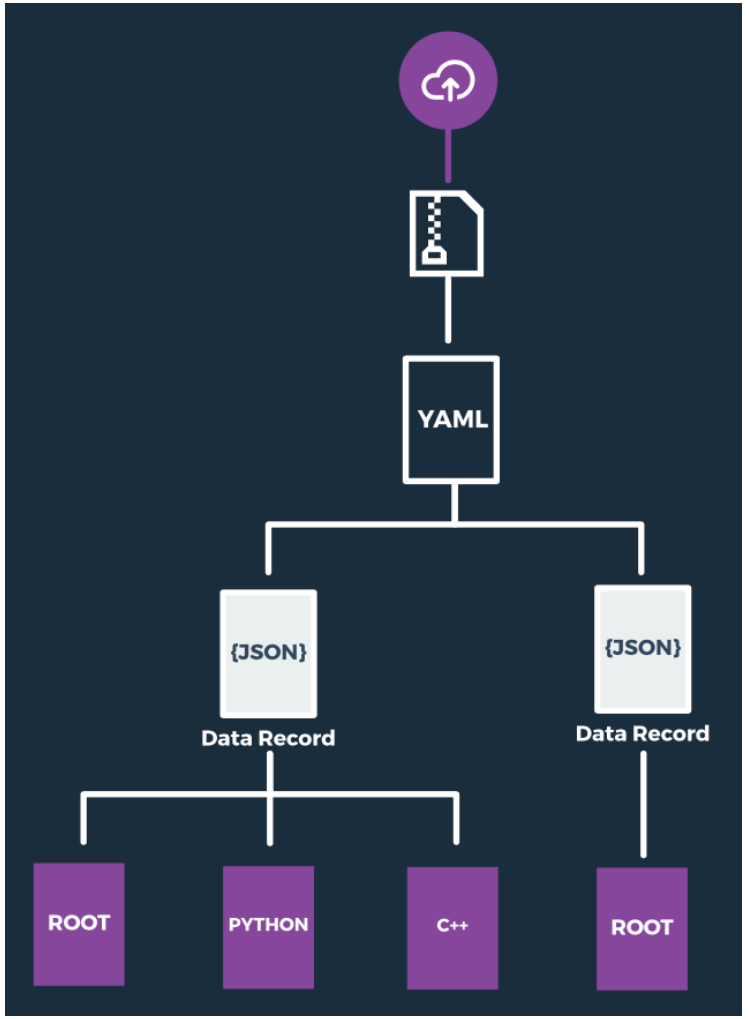
YAML (/ˈjæməl/, rhymes with *camel*[2]) was first proposed by Clark Evans in 2001,[10] who designed it together with Ingy döt Net[11] and Oren Ben-Kiki.[11] Originally YAML was said to mean *Yet Another Markup Language*,[12] referencing its purpose as a markup language with the yet another construct, but it was then repurposed as *YAML Ain't Markup Language*, a recursive acronym, to distinguish its purpose as data-oriented, rather than document markup.

Criticism [edit]

YAML has been criticized for its significant whitespace, confusing features, insecure defaults, and its complex and ambiguous specification.[32][33][34] Technically correct YAML can be completely unreadable by humans, especially if string values have leading or trailing whitespace.

Submission Tools

- YAML can be very verbose …
- Much has changed over the past 3 years
  - Conversion tools: migrate TXT/ROOT files into YAML format
    - Developed in CMS, see https://github.com/HEPData/hepdata_lib
  - (complicated) scripts that take input from web pages
    - Developed in H1 (DESY)
  - Validation tools: verify that the submission.yaml and YAML data files pre-submission



➢ Tools need some minimal adaptation to your results

- see https://github.com/HEPData/hepdata-submission

STAR's data repository

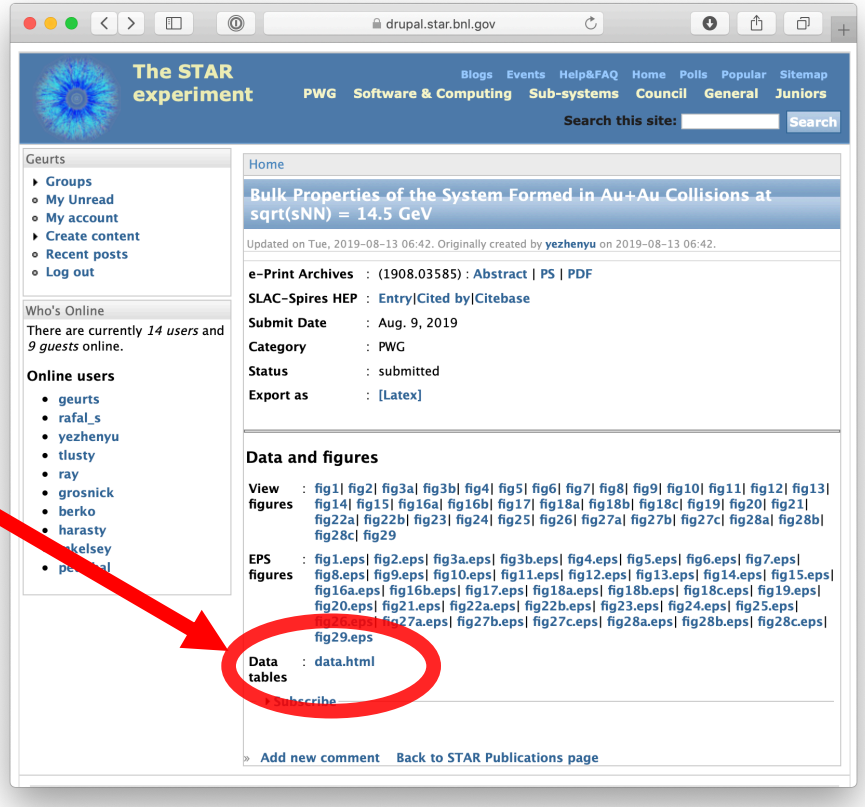
- Every submitted paper is accompanied with a data record
  - https://drupal.star.bnl.gov/STAR/publications

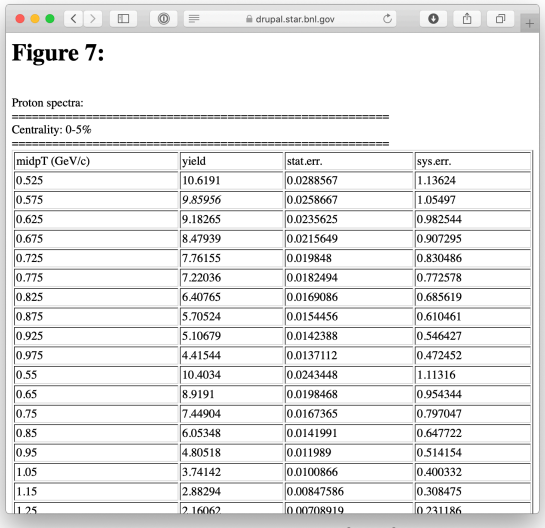
➢difficult to search

➢format mostly text-based, but less uniform

➢maintenance, versioning?

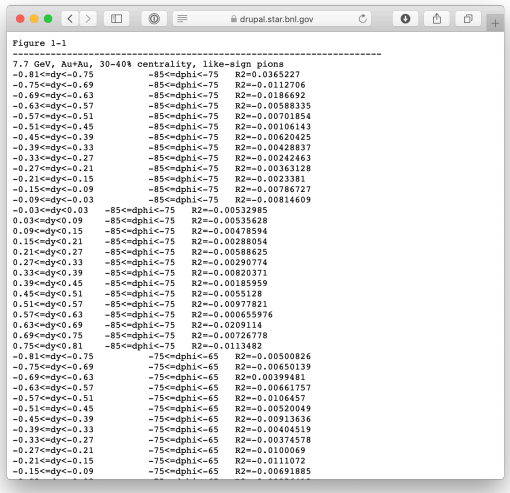
➢concerns about (long-term) access





1908.03585

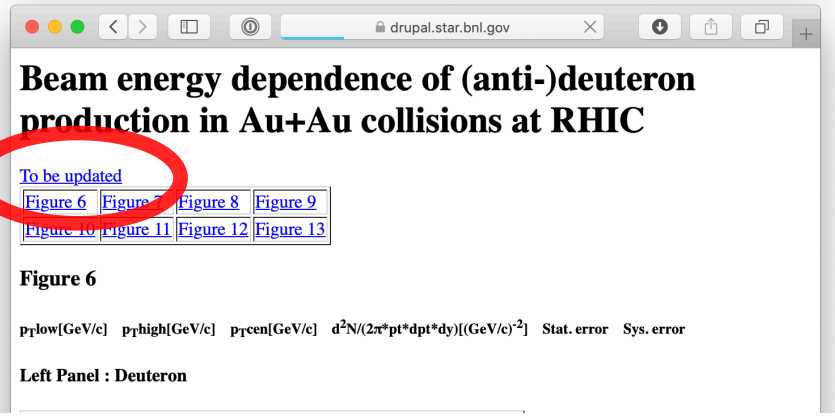
https://drupal [...] /316/data.html



1906.09204

https://drupal [...] /315/Fig/Fig1-1.txt

!!



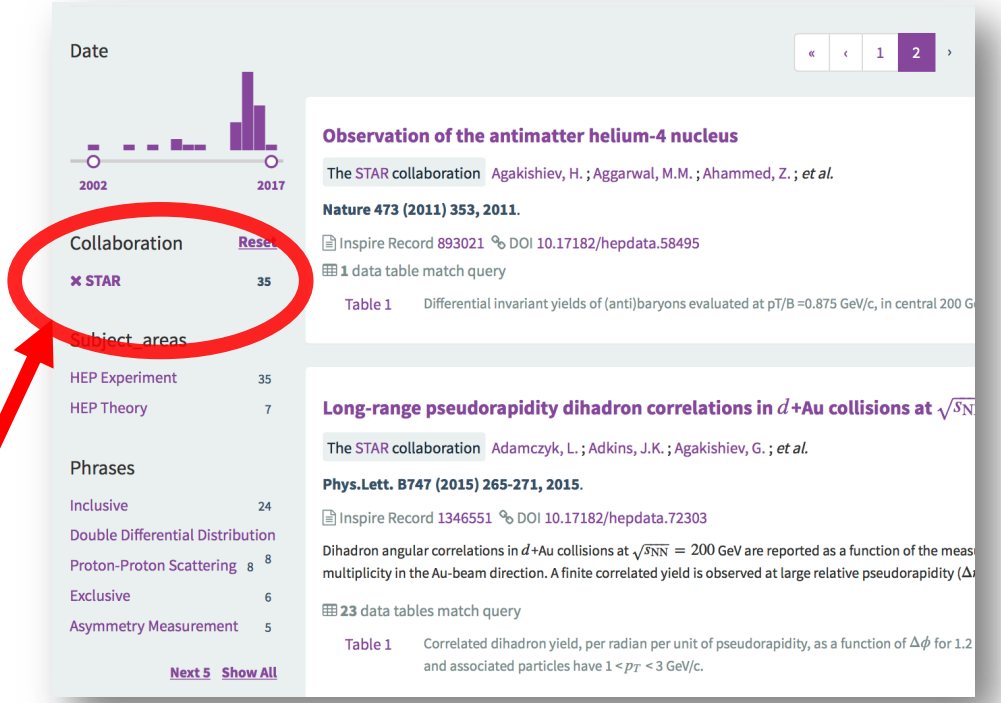
PRC 99 (2019) 64905

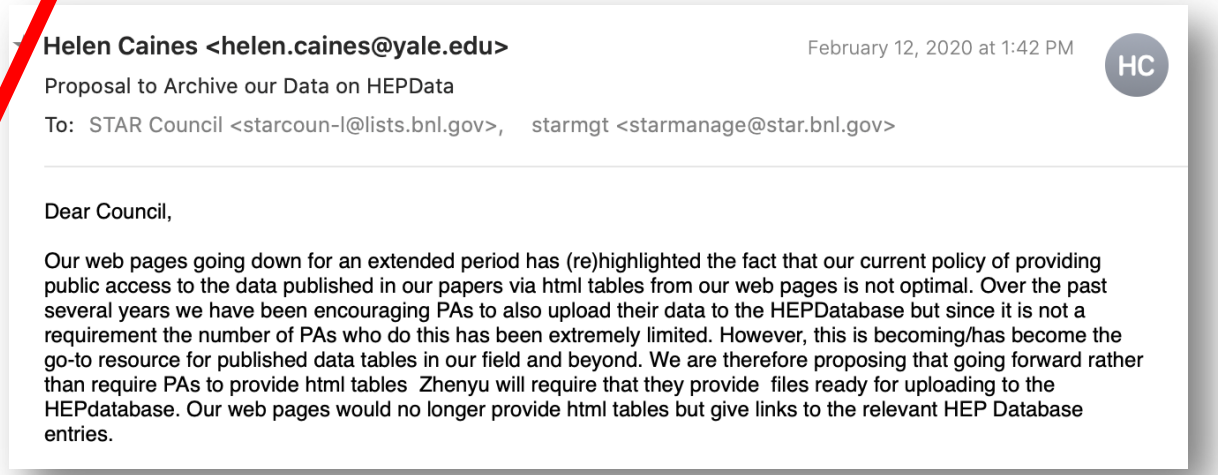
https://drupal [...] /302/data.html

STAR and HEPdata

- Before 2014: 7 papers in HEPdata
- 2016-2017 (from my Jan. 29, 2016 slides):
  - Nick Luttrell (Rice) is looking into defining simple templates that will help facilitate uploading STAR published data to hepdata.net
    - first test look encouraging (direct links between hepdata.net and inspirehep.net are very useful)
    - Past 184 papers? Still a challenge, but in the interest of a wider dissemination of our data worth the effort.
- Note: these tools were still based on

  OLD hepdata format
- May 17, 2017: 35 papers in HEPdata
- Feb.12, 2020: Proposal to Council





Proposal to Archive our Data on HEPData

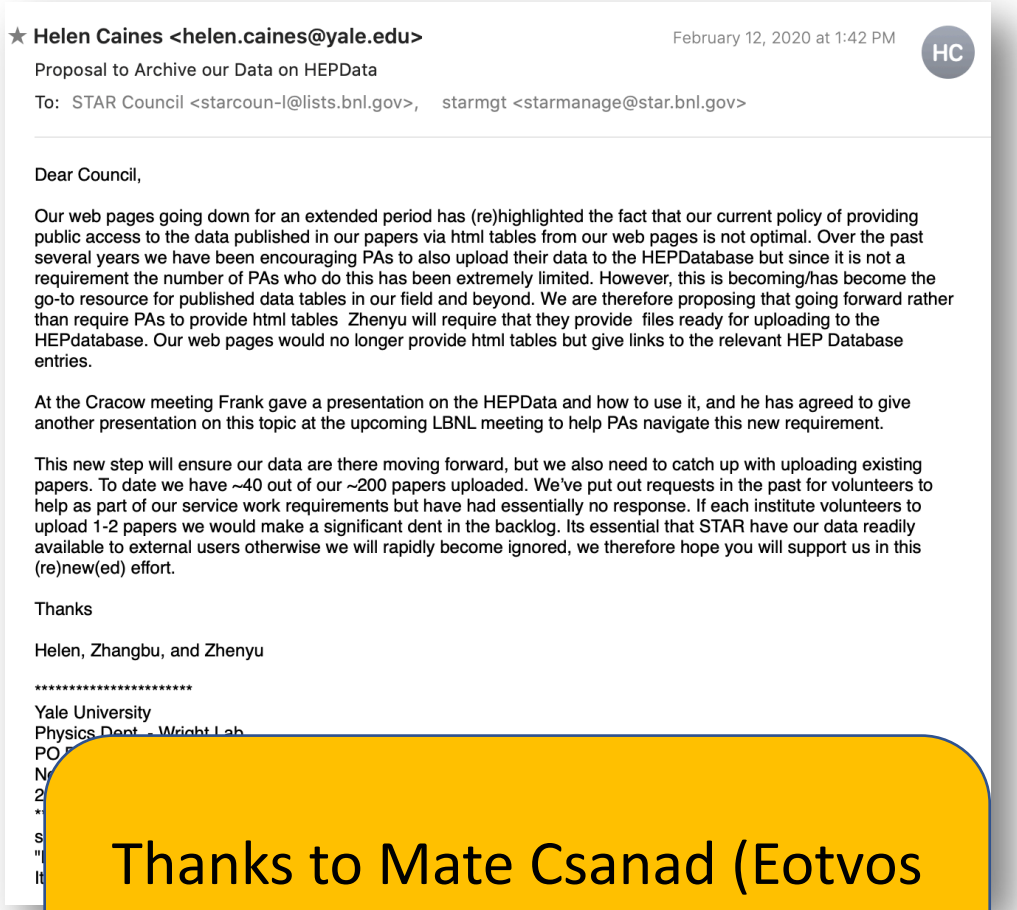
[email from Helen & Zhangbu to STAR Council]

- Concerns about public access to published STAR data
  - highlighted by our webpages being offline for an extended period
- Concerns about format and overall consistency

<u>Proposal:</u>

- ❖ Switch away from HTML tables to files that are ready for uploading to HEPData.
- ❖ Include QA and sign-off in the GPC process
- ❖ Involve collaboration to catch up with the backlog
  - opportunities for institutes
  - call for volunteers

➢ *… essential that STAR have our data readily available to external users otherwise we will rapidly become ignored*



★ **Helen Caines <helen.caines@yale.edu>** February 12, 2020 at 1:42 PM

Proposal to Archive our Data on HEPData

To: STAR Council <starcoun-l@lists.bnl.gov>, starmgt <starmanage@star.bnl.gov>

Dear Council,

Our web pages going down for an extended period has (re)highlighted the fact that our current policy of providing public access to the data published in our papers via html tables from our web pages is not optimal. Over the past several years we have been encouraging PAs to also upload their data to the HEPDatabase but since it is not a requirement the number of PAs who do this has been extremely limited. However, this is becoming/has become the go-to resource for published data tables in our field and beyond. We are therefore proposing that going forward rather than require PAs to provide html tables Zhenyu will require that they provide files ready for uploading to the HEPdatabase. Our web pages would no longer provide html tables but give links to the relevant HEP Database entries.

At the Cracow meeting Frank gave a presentation on the HEPData and how to use it, and he has agreed to give another presentation on this topic at the upcoming LBNL meeting to help PAs navigate this new requirement.

This new step will ensure our data are there moving forward, but we also need to catch up with uploading existing papers. To date we have ~40 out of our ~200 papers uploaded. We've put out requests in the past for volunteers to help as part of our service work requirements but have had essentially no response. If each institute volunteers to upload 1-2 papers we would make a significant dent in the backlog. Its essential that STAR have our data readily available to external users otherwise we will rapidly become ignored, we therefore hope you will support us in this (re)new(ed) effort.

Thanks

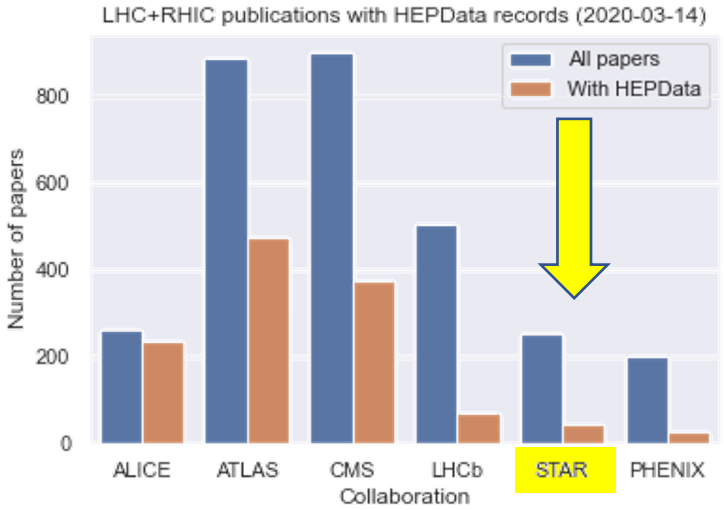
Helen, Zhangbu, and Zhenyu

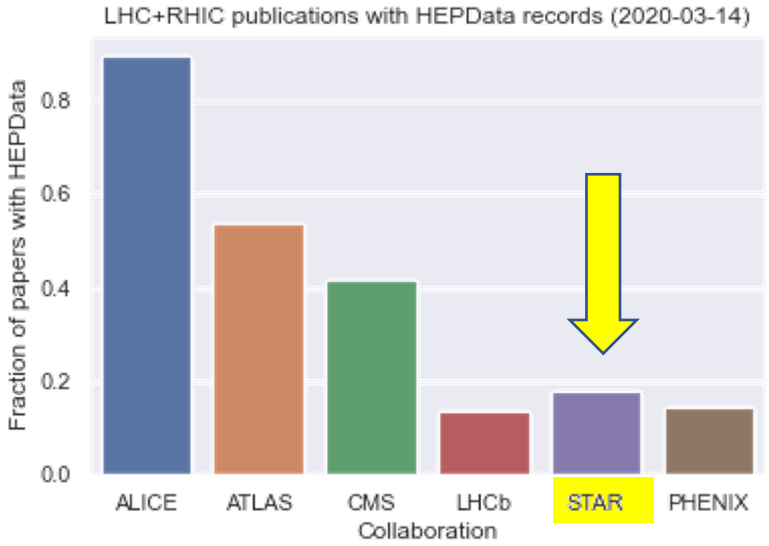
****

Yale University

Thanks to Mate Csanad (Eotvos Univ.) and Hao Qiu (Lanzhou) for stepping up and volunteer!

Where are we now?



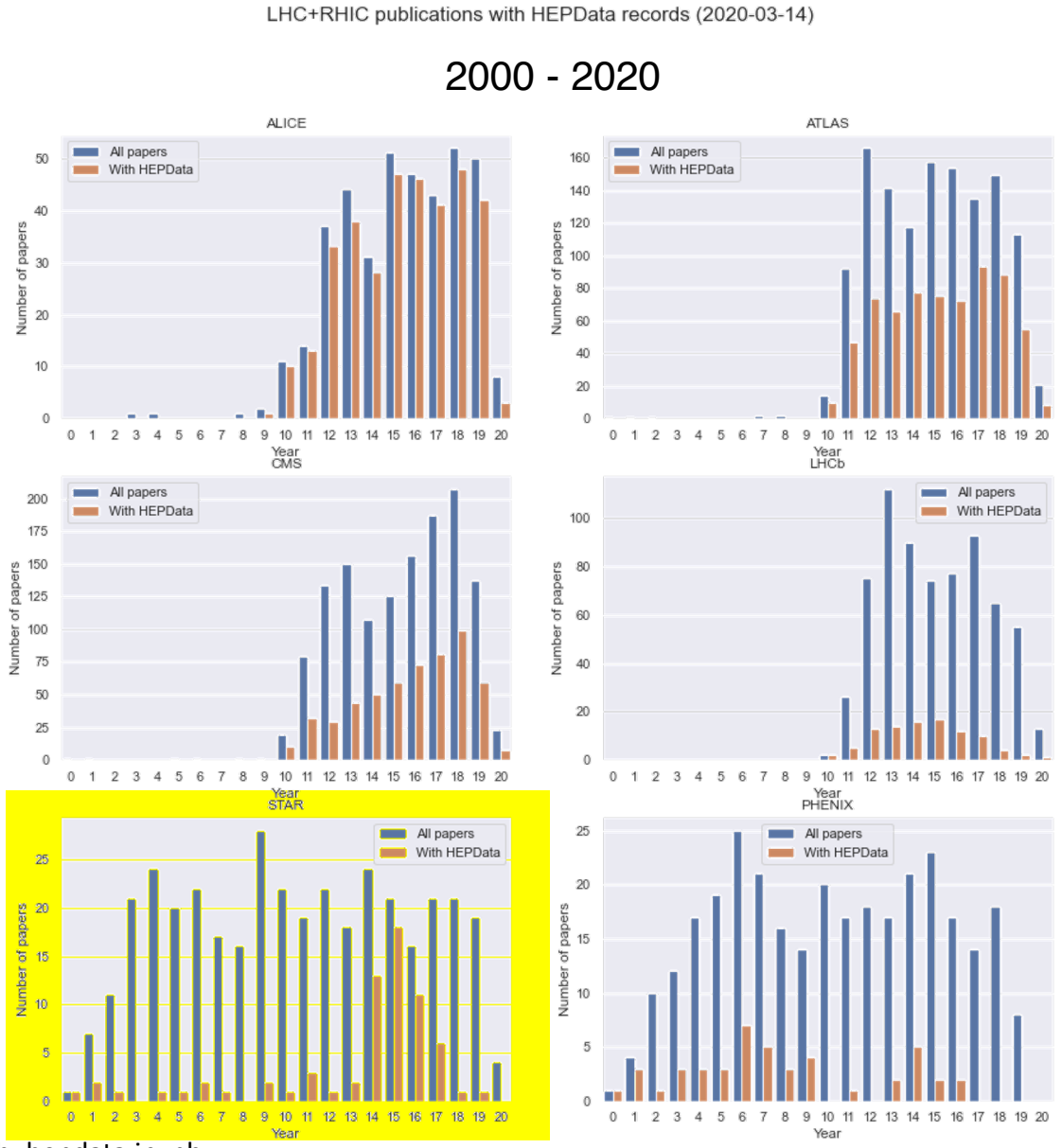


INSPIRE search query:

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

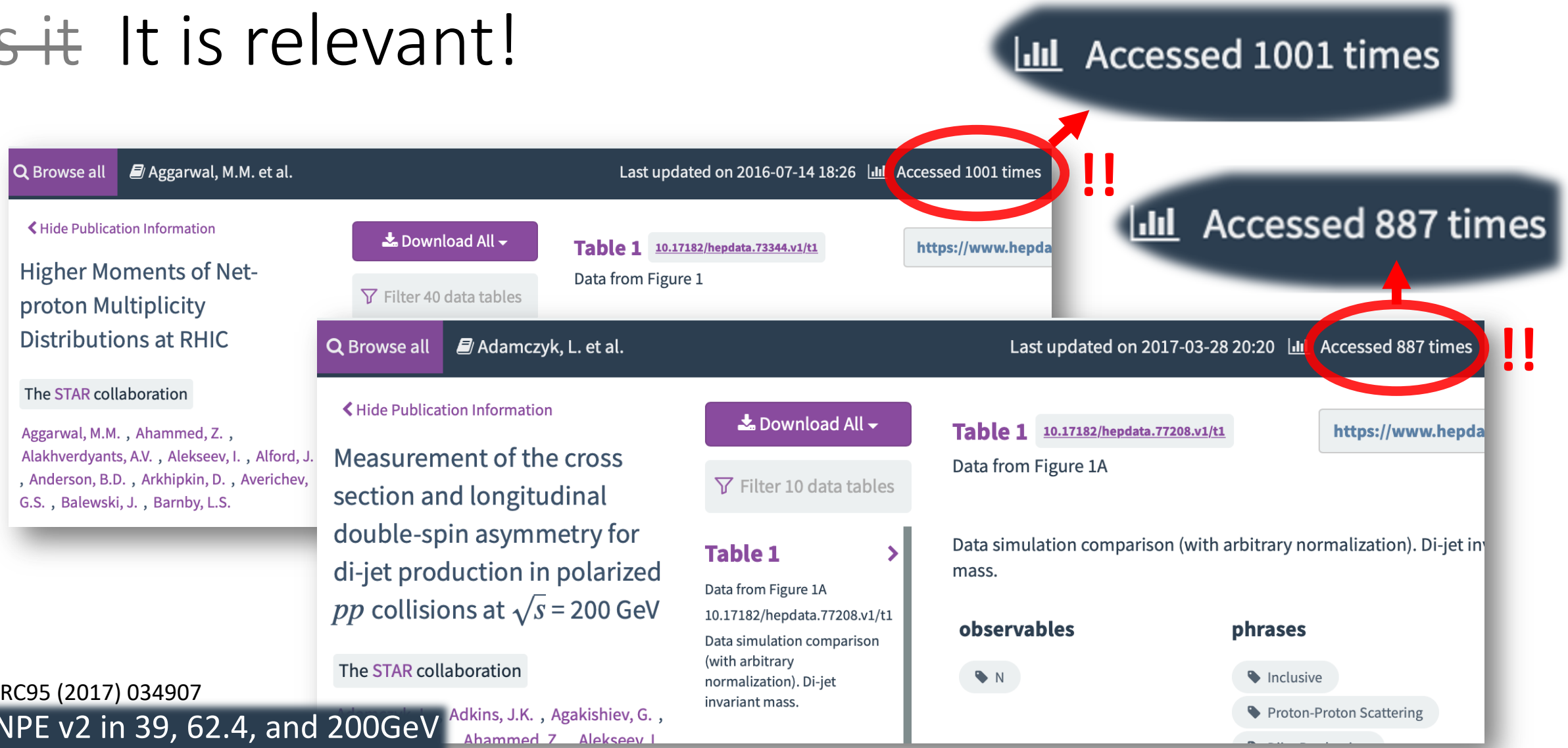
ALICE:	90%
ATLAS:	54%
CMS:	42%
STAR:	18%
PHENIX:	14%
LHCb:	13%

total sample size:
2989 papers (2000-2020)



Source: https://www.star.bnl.gov/protected/lfsupc/geurts/notebooks/hepdata/count_inspire_records_with_hepdata.ipynb
Based on https://github.com/HEPData/miscellaneous/

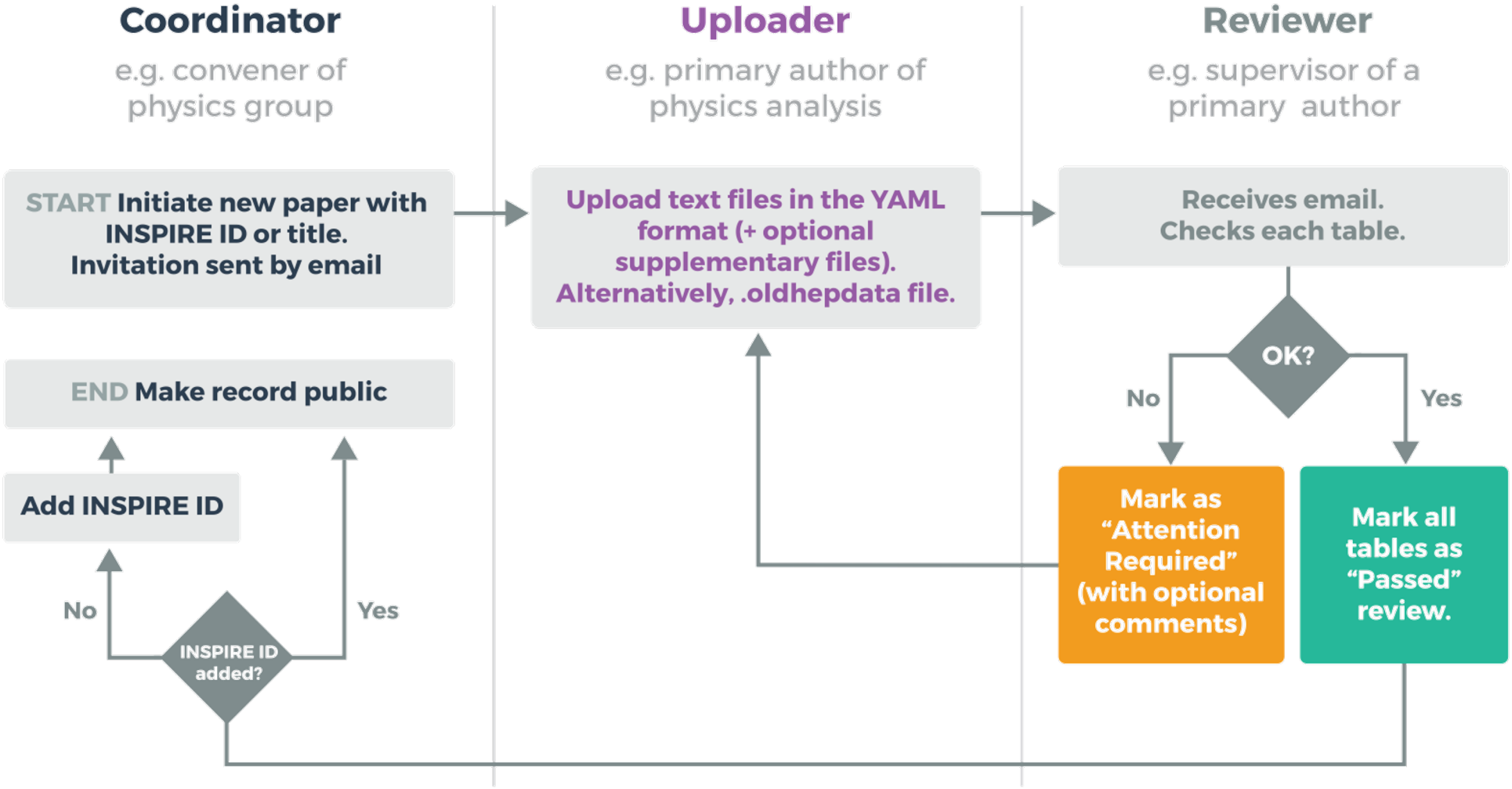
~~Is it~~ It is relevant!



PRC95 (2017) 034907

NPE v2 in 39, 62.4, and 200GeV
1008 times!!

HEPData Submission Protocol

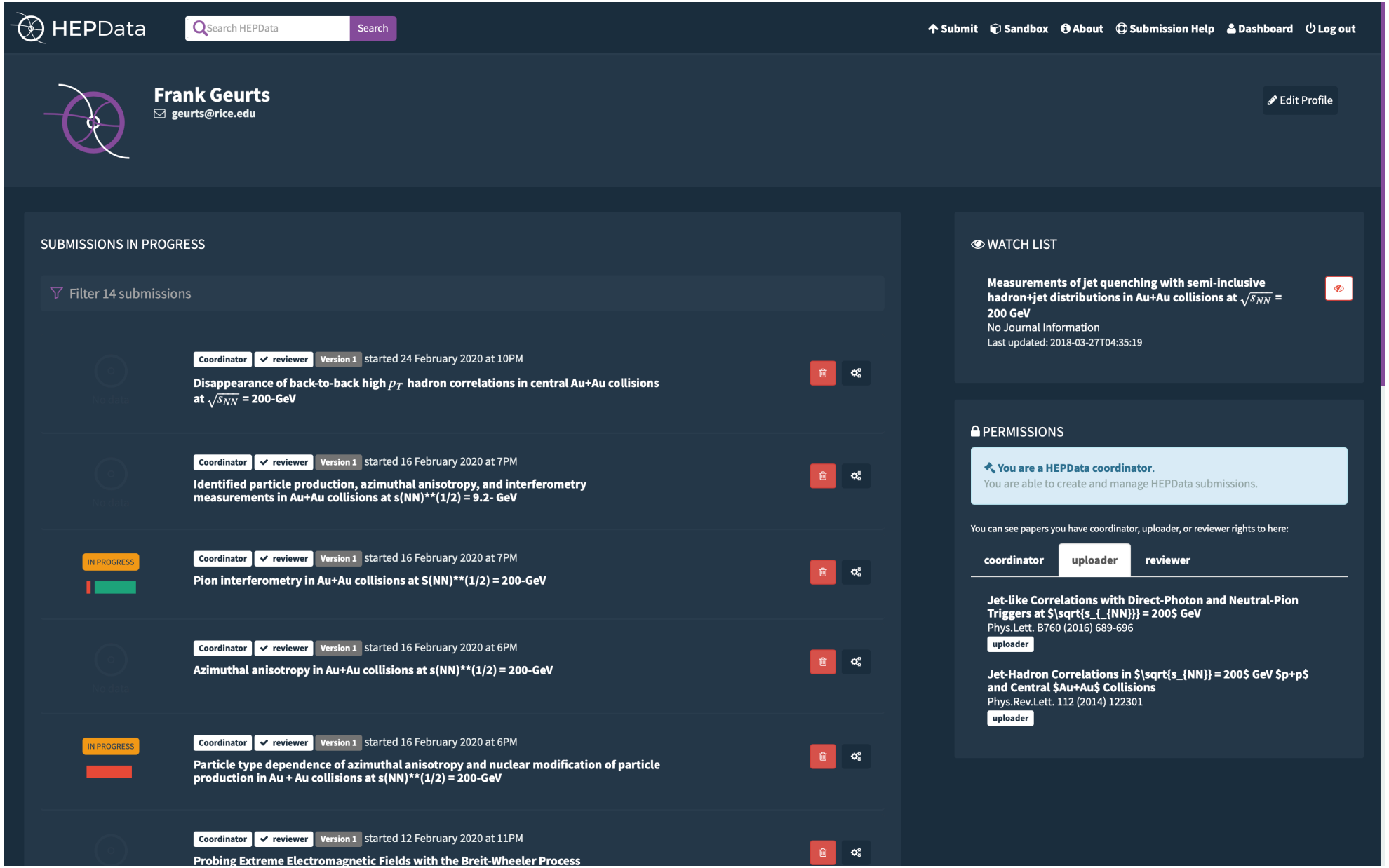


- Dedicated *dashboard* for each role: http://hepdata.net/dashboard
- *Sandbox* allowing to test and share uploads (persistent URL, record can be removed)

Coordinators
ALICE : 1
ATLAS : 7
CMS : 9
LHCb : 1
STAR : 1

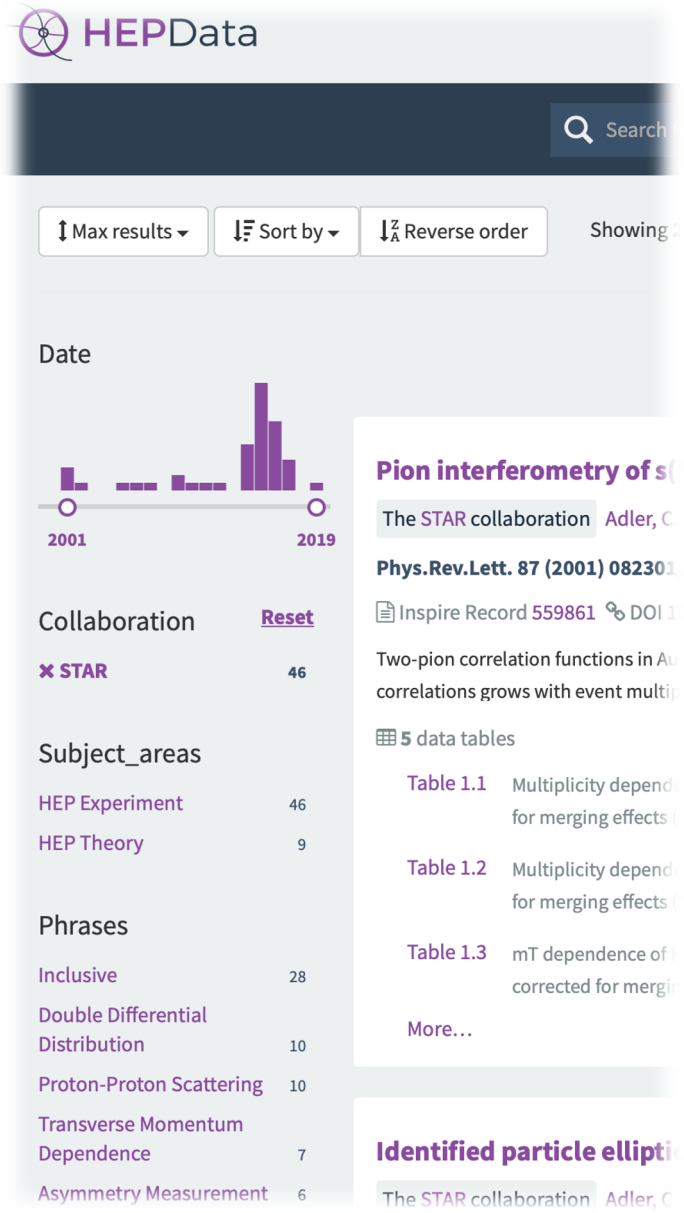
Maguire, arXiv:1704.05473

HEPData Dashboard



What is next for STAR?

https://www.hepdata.net/search/?collaboration=STAR



- Current status:
  - ✓46 papers submitted (40 in old HEPdata format, 6 in new HEPData)
    see https://github.com/RiceU-HeavyIons/STAR-HEPdata
  - ➢ **~200 papers still to go …**
- Proposal:
  1. reintroduce HEPData submission in STAR's GPC
     - PAs prepare appropriate YAML files
     - GPC chair/delegate signs off acts as the reviewer
     - after paper submitted to arXiv, STAR coordinator initiates HEPDb record
     - ➢ **PA uploads, GPC chair reviews, HEPDb coordinator finalizes**
  2. catch up with existing backlog
     - adapt existing tools (CMS, H1, etc.) to STAR's html/txt files
     - prepare and upload YAML tarballs
     - review new entries
     - ➢ **multiple STAR service tasks and/or volunteers (contact me!)**

More Resources

Look up STAR in HEPData

- https://www.hepdata.net/search/?collaboration=STAR

STAR Examples (mostly old format, but also a couple using the new format – thanks to Christine Nattrass)

- https://github.com/RiceU-HeavyIons/STAR-HEPdata

ALICE tutorial

- http://alice-publications.web.cern.ch/[...]/HOWTO_new_HEPData_authors_0.pdf

Documentation from HEPData team

- Recent conference paper: J. Phys.: Conf. Ser. 898 (2017) 102006 (arXiv:1704.05473)
- Documentation with more details/examples https://hepdata-submission.readthedocs.io/

Conversion tools

- https://github.com/HEPData/hepdata_lib