

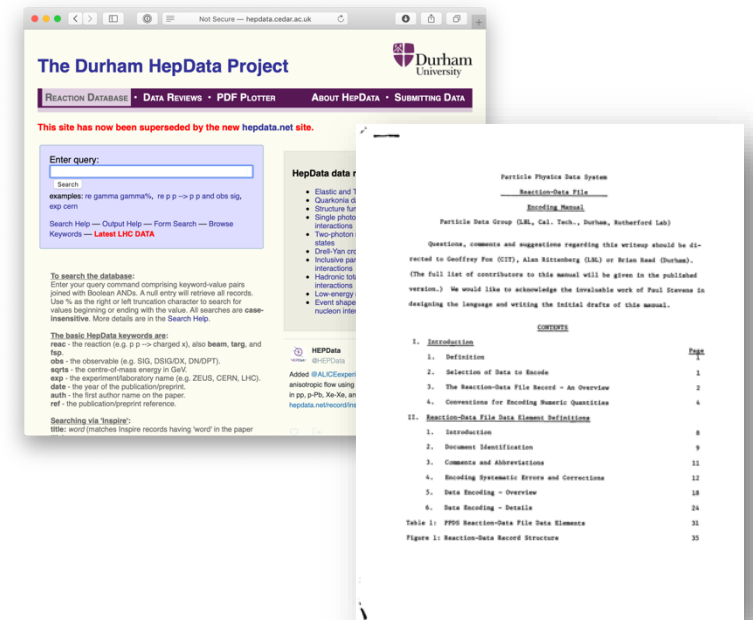


HEPData Feedback STAR Collaboration

Frank Geurts
Rice University

STAR in HEPData

- **Pre-2014** :: six STAR papers (old data format)
- **2016-2017**: reintroduce HEPData to collaboration
 - Rice student uploaded 28 key papers, all based on the old format
 - Initial feedback: *need for templates and improve uniformity of submissions, e.g. specific metadata appropriate to heavy-ion physics*
 - Informal survey (2016) among a couple of LHC experiments: *no explicit collaboration-specific instructions*
 - by 2019, 40 papers were uploaded in the old format
- **2017/2018** :: first proposal to **integrate** HEPData submission into STAR's paper preparation workflow
 - PAs and GPC Chair take on formal roles as uploader and reviewer
 - PAC and HEPData coordinator initiate and finalize
- **2019** :: prepare collaboration
 - concerns about tools, formatting, uniformity/consistency



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.

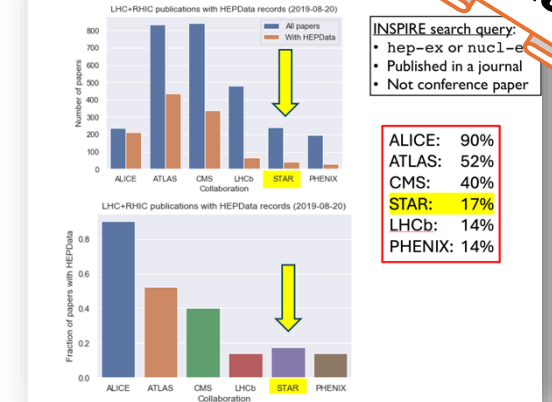
Criticism [edit]

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.

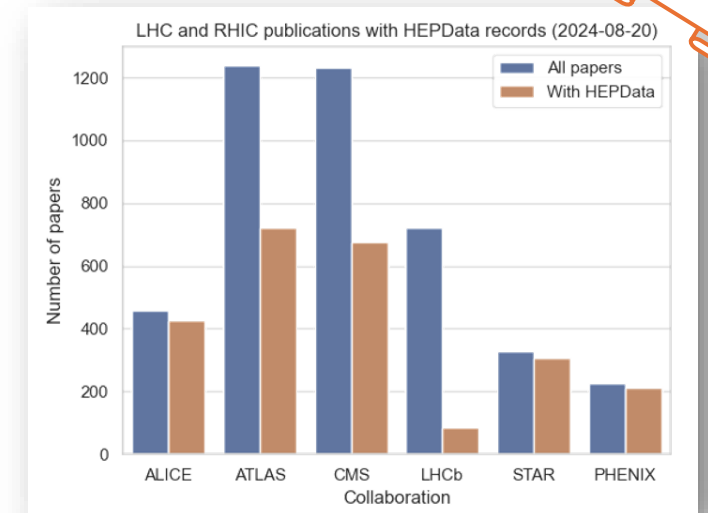
Formal Adoption by STAR

- **March 2020** :: formally adopted by STAR Council for all new paper submissions
 - HEPData prime location for STAR published results
 - collaboration-wide initiative to tackle very large backlog
- **STAR HEPData Shifts** — Aug. 31, 2020 – July 11, 2021
 - shift coordinator: Mate Csanad
 - shift credit spanned three RHIC Runs (20-22)
 - 180-200 papers in 73 double shifts
 - data digitalization where txt/root files were not available
 - average shift: 3 weeks for 3 papers (~20 tables)
 - **fun facts**: most #tables per shift: **34**; most papers per shift: **6**; most shifts by single person: **4 (12 papers)**; most shifts by single institute: **8**
- **Now** :: 321 papers uploaded in HEPData
 - another ~14 submissions pending journal review

Where are we now?



2024



Feature Requests

- **Coordinator Tools**

- Quick access to statistics:
 - Active uploads (dashboard can improve, e.g. show uploader/reviewer names on the same page)
 - Spreadsheet with all relevant entries, reviewer, uploader, #access (vs. time), #tables, #date, revision#, DOI, inspire, etc.
- Address Book functionality
 - access to past submitters/reviewers
 - more flexibility in associated email accounts (ORCID)
- Approval fine-tuning
 - allow individual tables to be re-approved without the need to reapprove all

- **Post-submission metadata/key-words**

- Keywords may need to be added at later stages
- Highly relevant for data-preservation efforts

- **Improving control over consistent submission format**

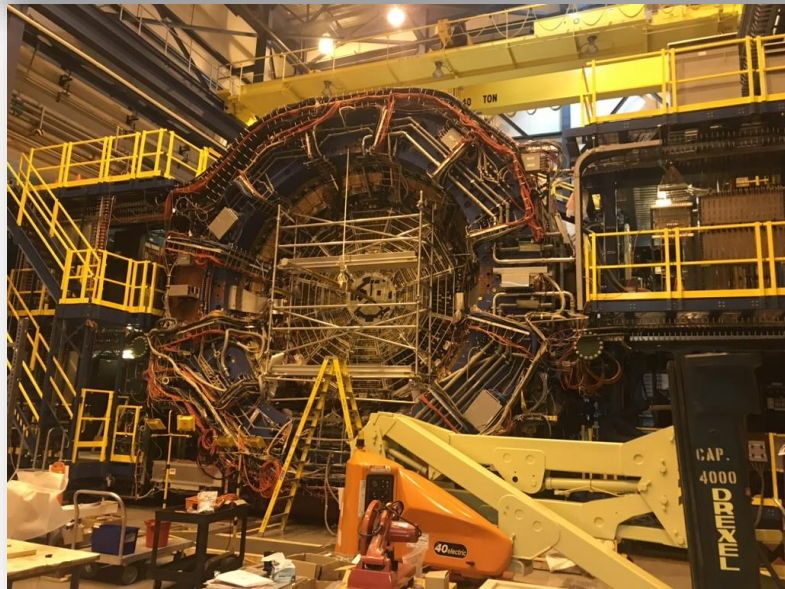
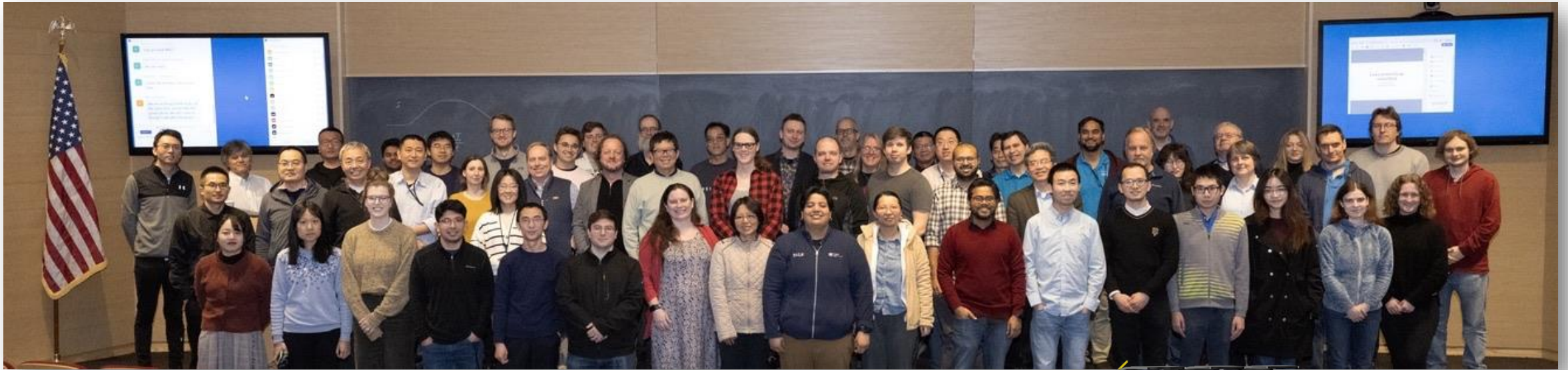
- e.g. templates that better facilitate agreed-upon collaboration standards

- **Handling of uncertainties** (asymmetric, etc.), **X-Y plots**

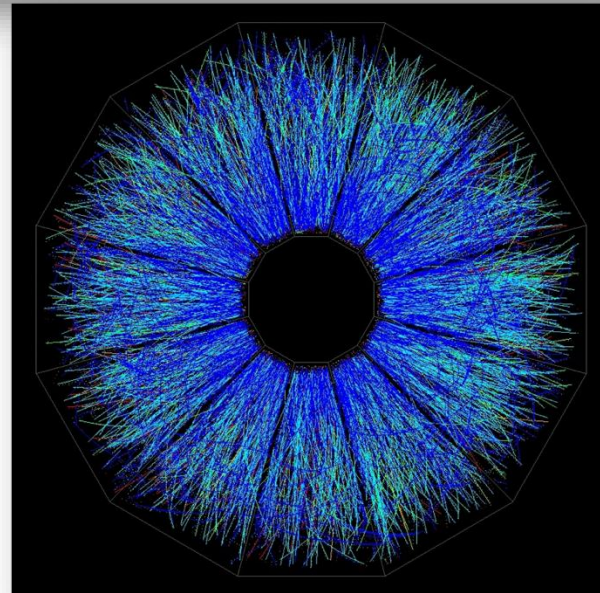
- **Handling of old - pre-YAML - submissions**

- assign to the current coordinator

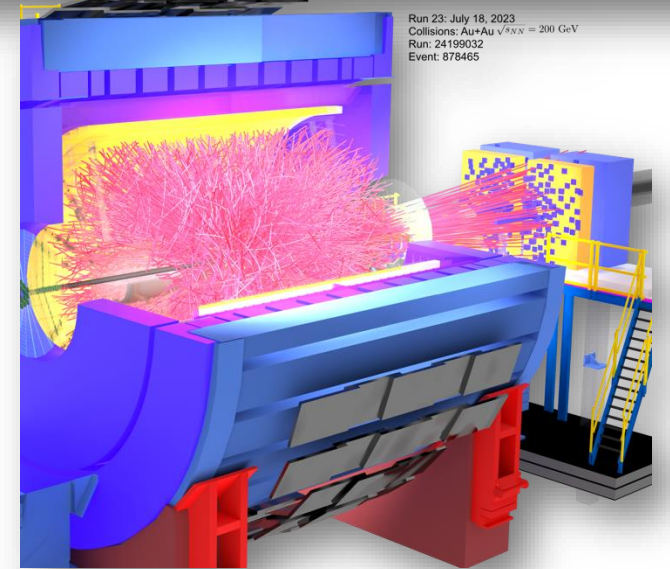
On behalf of STAR: Thank you!



1/15/25



HEPData Users' Meeting 2025



5