Hepdata:
My ATLAS perspective

Bill Murray
STFC-RAL / University of Warwick
bill.murray@cern.ch
28th January 2020

**Thanks especially
to Louie Corpe,
UCL**

- **How important is HepData?**
- **What could be better?**

Background scene

- Collider physics is at a crossroads: since discovery of H (and non-discovery of much else) we don't have a guiding theory anymore.
- This means a paradigm shift from theory-driven (top-down) to data-driven (bottom-up) approach to locating new physics, if it is accessible. This means we don't know what we are looking for, and worse, we need to sift through an unprecedentedly-large dataset to find it.
  - unlike in the past when we had a fairly good idea
- So it's absolutely crucial that all measurements we make are set up to be re-used/re-interpreted as new ideas come out from the theory community
  - And we need to be able to correlate between measurements in case this provides clues.
- Furthermore, the size of the LHC dataset now means that many analyses are no longer stats-limited. Systematics are much more likely to be correlated between bins. So it's essential we can keep track of that too.

Hepdata deepening

- In order to fully exploit the LHC data, and the HL-LHC data, we need a robust database of all the measurements/searches which are made.
- It needs to be easy to navigate and access information, and it needs to contain all the info we'll need to re-interpret/re-use a search/measurement
- HEPdata (+Rivet) provides this.
  - If it didn't exist, we'd need to invent it.
  - It's vital to being able to fully exploit the potential of the LHC and HL-LHC.
- There has been recent progress on improving the quality of HEPData records, and what's more to agree some standards between the experiments

Hepdata today

- HEPData has emerged as the industry standard database. This is in part because of it's close relationship with Rivet:
  - HEPData reference data is the reference data for Rivet routines.
- Since 2018, HEPdata-Rivet has a format for propagating uncertainty breakdowns from HEPdata entries into the reference files for Rivet. This is a huge step because now one can in principle evaluate bin-bin correlations when using Rivet. (see slide 16 of Graeme's May 2019 talk)
- Tools are in development which make fast scans of new models over the bank of LHC measurements (and now also searches since Rivet's smearing capabilities came online!)
- That means you can (approximately) check in O(hours/days) whether some new model is already excluded by the LHC.
  - CONTUR is a good example of this,
  - And uses the correlation info provided in HEPData
- This is only possible thanks to HEPdata's existence
- In summary, HEPData is at the forefront of the re-interpretation effort.

Feature requests

- Can we find a way to link tables/objects on HEPdata? i.e. binX of tableX is the covariance matrix for binY of tableY
- Can we find a way to make the upload/review of HEPdata entries less painful: even for experienced uploaders, it can be tricky
- Is the current upload interface scalable? The LHCEWWG recommendations would lead to a lot more info being uploaded, and the worry is that this could overload the system.
- Maybe there could be some link to gitlab-like interface where draft entries could be amended file-by-file rather than having to re-upload the whole HEPdata entry every time a small change is needed?
- More integrated support for additional objects:
  - e.g. Bootstrap Replicas, pyhf likelihoods, XML table for BDTs etc

Louie Corpe

ATLAS Generator infrastructure and tools convener

Central Hepdata role

- HEPdata is at the very core of data and analysis preservation and as such it is absolutely crucial if we want to be able to get the most of our data.
  - This is true not just for the LHC experiments, but modern particle physics in general, I would argue. We simply can't do without it.
- With the ever increasing precision, complexity and detail of our analyses comes the need to preserve more complex results too — we no longer just upload simple tables of numerical values, but have started to preserve full-blown likelihoods, BDT classifiers, multi-dimensional folding matrices. More data types will likely follow and the current object types native to HEPData have already started to show signs of insufficiency (consider the memory demands associated with large multi-dimensional folding matrices!), making it essential HEPData be funded and maintained properly if we want it to scale with the needs of the field for the foreseeable future.

Chris Gutschow
Physics Modeling working Group convenor

HEPdata does respond:

- I would like to highlight that recently HepData has been quite helpful in accommodating new types of data products on HEPData, especially the likelihood integration, for which we can now have individual permanent identifiers / DOI, which was previously impossible.
- To be honest, it took a bit of time to get there, partly due to the transition to the new system, but I think it's an encouraging sign for the future development of HEPData in which a richer set of data can be stored and retrieved..

Lukas Heinrich

ATLAS Analysis model group coordinator

Top Working group

- We would like to express our high appreciation for the HEPdata service.
  - Making our measurements available in electronic form is an important part of the publication process.
  - We wish our records on HEPdata would be more complete.
  - Many analysis teams react with a considerable inertia on our requests. They are exhausted after the long publication process in ATLAS.
  - In many cases the main analysers (most of the time PhD students) graduate after WG approval, even during the publication process, and it is difficult for teams to find people to work on HEPdata.
  - We see this as the main challenge for our group.
- One could certainly argue about the format (YAML) being a bit awkward, but this is not the main point.

Francesco Spano and Wolfgang Wagner
Top Working group conveners

SUSY working group I

- We have compiled a list of suggestions/comments concerning HEPData:
  - Stability issues: The down-time of the website is noticeable.
  - A clear and updated list of tags (processes, observables, etc.) to be used would be very helpful
    - https://github.com/HEPData/hepdata/issues/60
  - Version control for uploaders submissions would be helpful.
    - SUSY group have built their own, but better if HEPData did
- Suggest an 'Accept All entries' button for reviewers. Going through a lengthy record N-th time entry by entry just to have them ticked is really tedious.
- The plotting macro should make a plot even if negative or zero entries are present rather than not plotting anything at all as it is currently done -- https://github.com/HEPData/hepdata/issues/131
- Possibility to download complete, with plots etc desirable
  - https://github.com/HEPData/hepdata/issues/114

SUSY working group II

- The conveners wish to express their enormous appreciation and thank you for the continued support, financial and otherwise, of HEPdata, which of course plays a vital role in our community.
- It may also be worth mentioning that we are now publishing full likelihood information in our SUSY HEPdata records (pub note here https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-029/

  and example in HEPdata here: https://www.hepdata.net/record/ins1765529

Federico Meloni and Laura Jeanty
ATLAS SUSY conveners

ATLAS SM group

- SM group extensively use HEPDATA (and Rivet) as part of their data-preservation strategy
- The SM group produces precision measurements that probe the strong and electroweak forces.
- All measurements are now required to submit a HEPDATA record, which contains the measured data and the best available theoretical predictions.
- They work closely with theorists and other experimental groups, within the LHC EW WG, to identify the best data-preservation strategy for the future
  - such as linked statistical and systematic correlation matrices
  - **This is likely to lead to requests for additional functionality in the future.**
- They therefore strongly support the continuation and development of HEPData as a tool for the high-energy physics community.

ATLAS SM group request

- The 'approval' of the HEPDATA records requires each individual plot to be signed off.
- This does not match well with the workflow,
  - Records are downloaded en-masse
    Algorithms used to assess whether they match the results published in papers.
- It would be good to be able to sign off all plots at once.
  - Possibly this functionality exists, but we are unaware

Andy Pilkington and Evelin Meoni
ATLAS Standard Model Conveners

Conclusions

- This is not my talk
  - It is a collection of inputs from ATLAS physics coordination
  - All input received is reported, (sometimes slimmed)
- I am to be replaced: Richard Hawkings has agreed to take this role
- HEPData is an integral part of ATLAS Physics output
- There is a continual series of request for incremental improvements

"We simply can't do without it."