CMS & HEPData

Henning Flaecher
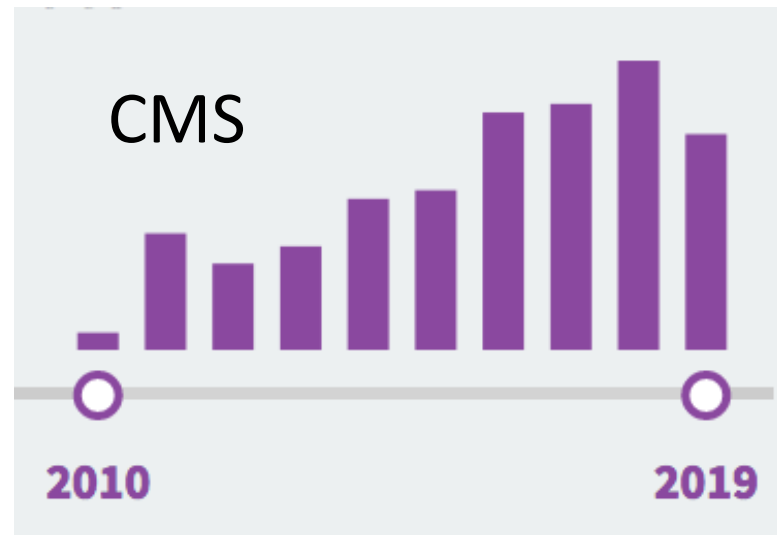
with feedback from some CMS HEPData
PAG contacts

(A. Albrecht, C. Lange, R. Mankel)

HEPData usage by CMS

- **All analysis results from published papers should be available via HEPData**
  - **CMS Collaboration Board (in 2016/17?)**
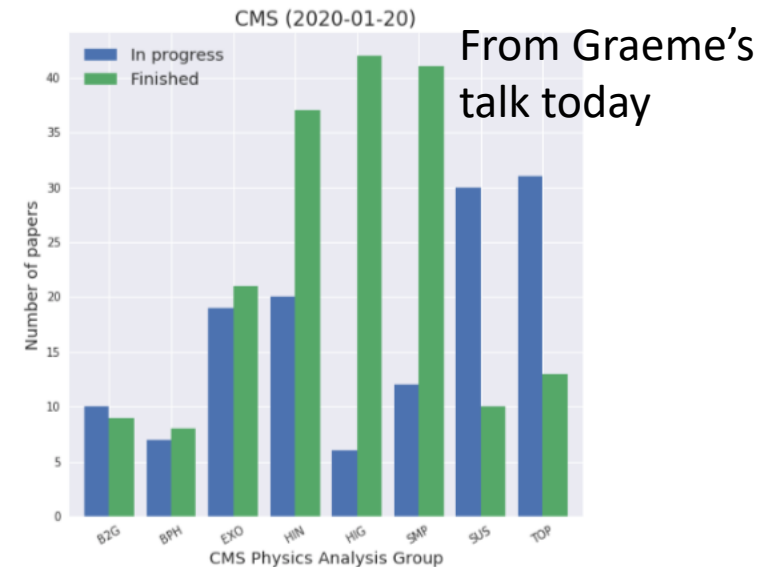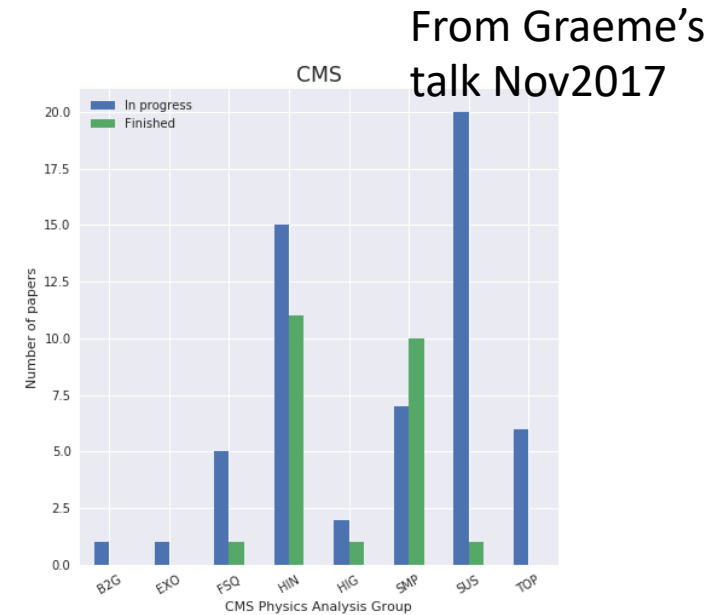
CMS

2010   2019

- Current status:
  - 2019: 377 papers on HEPData out of 949 collider data papers (40%)
  - Status of Nov 2017: 220/648 submitted Physics papers, i.e., 34%
  - 31% in 2015

HEPData usage across CMS

- All PAGs are using HEPData
  - "compliance" depends on physics analysis group
  - For SM physics groups usage is rather good
  - Historically less so for searches in SUSY and EXO
  - but many in progress
  - Top -> RIVET
- Link to HEPData entry from CMS publication pages

437 users created an account

From Graeme's talk today

HEPData "automated" submission

- "New" upload process in place for a while now
  - considered a big improvement, in practice for CMS:

- Coordinators and reviewer roles handled by HEPData Physics Analysis Group contacts (and physics group convenors)
- Uploaders are typically the analysis authors

- Some tools developed in hepdata_lib to facilitate conversion of plots and tables to YAML
(C. Lange, A. Albert – B2G,EXO)
  - https://github.com/HEPData/hepdata_lib

- It still often takes several iterations between reviewers and uploaders

Comments/feedback on analyst's side

Submission process has improved but is still considerable effort on analyst's side:

- HEPData and its format often not considered by analysts from the start
    - Conversion of tables/plots into YAML
    - Automated conversion of LaTeX tables?
        - sth for hepdata_lib?
    - Questions about clarity and where to find documentation
    - "Once we figured how to structure everything in YAML, it was fairly straightforward converting tables and plots"

- psychological/attitude problem: HEPData upload comes at the very end of a tiring approval and publication/review process
    - "not this as well…"

Comments/feedback on reviewer's side

- Sometimes "clunky" approval, small changes require reapproval of all tables, some pubs have lots...
  - Can all tables be approved in one go?
- Version history and diff feature for unpublished versions?
- Can comments from previous versions be retained?
  - E.g. keep a table with all comments and associate a "version number" and option to "approve" (but not delete)
- Sharable link for unpublished entries for review in a wider group?
  - Currently reuploading to user sandbox which is cumbersome
- General support/discussion forum for uploaders?
  - Most questions go via PAG contacts

Additional feedback from HEPData contacts

"Wishlist" for improved Performance/Features:

- Could rendering of tables be made optional?
  - slow performance for very large tables
- Is it possible to automize adding journal information?
  - not available at time of submission
- Scope to improve mapping of papers between CMS internal identifier to inspireID?
  - e.g. offer free text field for reviewer's use that displays on Dashboard
- Option to use arXiv identifier in addition to inspireID for creation of new record?
- Could list of observables be reviewed/updated (as well as phrases and particles) to have better defaults for physics processes?
  - consistency desired as otherwise people will come up with their own names
  - fix "features" for consistency, e.g. Cmenergies 13000 vs 13000.0
- generalised format for (simplified or full) likelihood preservation
- Improved documentation, some links still refer to "old" HEPData portal

Conclusion

- HEPData plays and important role and is essential tool for CMS analysis preservation and dissemination of results
  - strongly support its continuation
- Upload procedures and contact persons in place in all CMS PAGs
- Desire to further simplify/automatize the upload and review process
- Desire for further standardization/update of physics processes/observables etc.