



CMS feedback and hepdata_lib

Clemens Lange (Paul Scherrer Institute PSI)
HEPData users' meeting, 15th January 2025

Creating a HEPData record is mandatory for each CMS analysis

- > This policy became effective at the end of 2020
- > Dedicated HEPData contacts iterate with the analysts, the publication committee triggers the process
- > The DOI of the record needs to be added to the manuscript before submission to arXiv

What changed?

- > Andreas Albert and I managed to convince the collaboration that “HEPData is easy”
- > We developed [hepdata_lib](#) with documentation and examples and trained others on using it

However, creating a HEPData record can still be made easier

Recently, CMS has started to release full statistical models, now linked from HEPData

- > Model files can be too big for HEPData

Effort required to create a HEPData record



Based on a small survey (8 responses) and some personal discussions:

- > Most analysts can create a record within a single full working day
- > Some people, however, require several days (up to a week)

Issues:

- > People typically only create a record every other year if at all — examples and extensive documentation are important
- > For some it is difficult to create the input for HEPData
- > **Metadata, i.e. keywords, are not sufficiently standardised/unclear/too general**
 - Is there any monitoring on whether visitors of the HEPData portal actually use them?
 - Physics process description probably biggest issue, a lot of “new physics” hard to describe using HEPData keywords

1. Make it possible to have several (at least 3?) uploaders for one entry, with equal rights, that can be entered both from entry creation screen and entry modification screen
2. The same could be also helpful for reviewers
3. For review, when a new record is uploaded, only mark Tables for review again if they have changed
4. Allow editing of record title (before it is taken from INSPIREHEP)
5. Make "notify uploader and reviewer" the default action, (checked mark) with the possibility to uncheck it
6. Make it possible to share the entry link for viewing before it is finalised (e.g. to share with ARC chair or PubComm), i.e. share with a viewer token. Currently AFAIK people use sandbox entries for that, rarely add ARC chair as an extra reviewer (or even email zip archives).
7. For generator fragments and code snippets, allow rendering them inline rather than having to download them

Improving and maintaining hepdata_lib



Andreas Albert left HEP a few years ago

- > Since then, HEPData team (Graeme, Jordan) have provided help with maintenance and code review — thank you!
- > Thank you also to other contributors who have provided new features and fixes!

I can only maintain the library with relatively little effort thanks to **extensive tests and automation**

Few new features added overall

- > Library covers the most commonly used needs

There are still quite a few issues/feature requests

- > See https://github.com/HEPData/hepdata_lib/issues
- > This is certainly not a complete list — my experience is that most users will not create GitHub issues but rather somehow make things work for them (even if it takes them much longer)
- > Any help getting those open items addressed would be highly appreciated!

