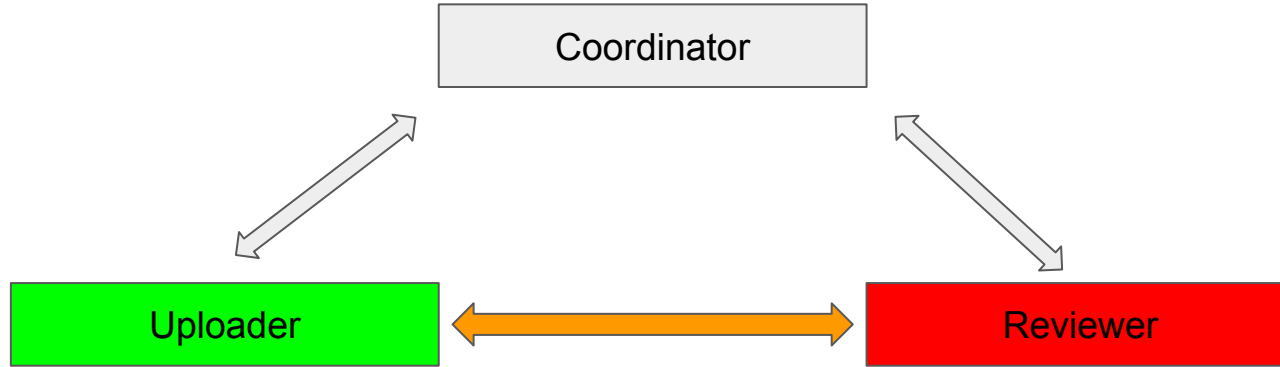# The PHENIX DAP update

- HEPData procedures - the upload and review process

Maxim Potekhin
(BNL, NPPS)
12/21/2020
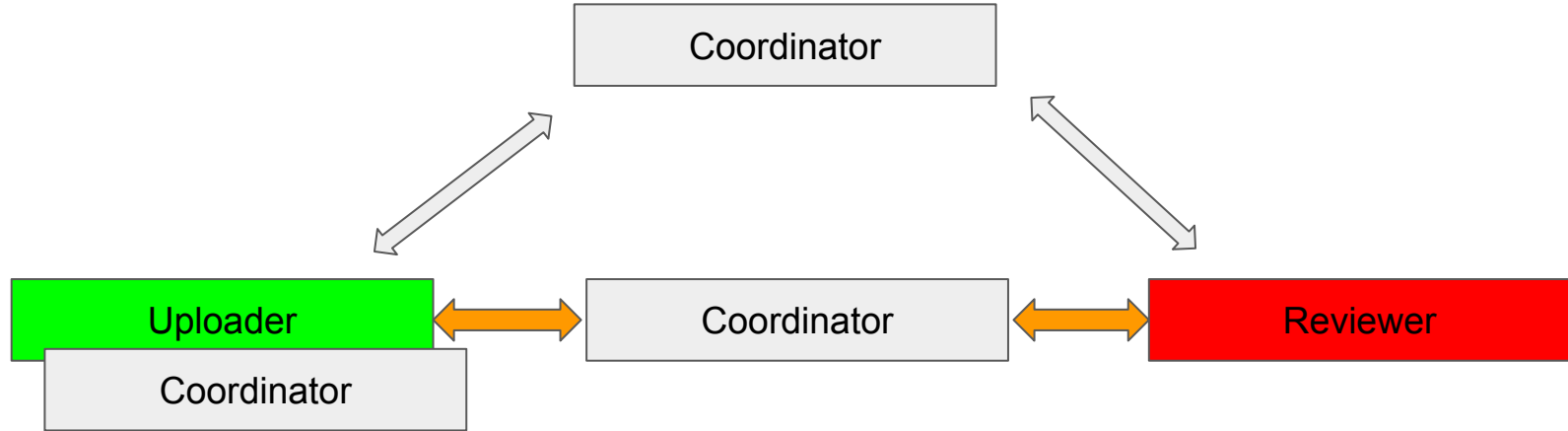
The HEPData workflow

- Someone (e.g. a member of a PPG) prepares a submission package and tests it with the "sandbox" feature on HEPData; materials are on GitHub
- Maxim creates an entry which is tagged by the InspireHEP ID, which for now is just a placeholder; the "reviewer" and the "uploader" are specified at this time
- The uploader uses the GitHub content to create a tarball for HEPData submission
- The uploader uploads the materials (until now this was usually Maxim)
- The reviewer reviews every item within the submission, and (optionally) leaves comments
- The uploader makes fixes and re-uploads, repeating the loop
- When all is checked and considered OK, Maxim finalizes the item
- It becomes public

The HEPData workflow - canonical

The HEPData workflow - current process

The problem

- If a submission is 100% correct the review process is just a formality
  - So the coordinator can do it all - this was what we did in Summer 2020
- As we started the HEPData effort in earnest this is often not the case
  - Complex submissions with many elements are created
  - Errors of all kinds are unavoidable
- If the upload duty is assigned to the coordinator he needs to relay all the information from the reviewer to the uploader for corrections
  - This can be a lot of info
- The HEPData site is designed to support direct communication between the uploader and the reviewer as "chat" on the webpage of the submission
- If the coordinator is in the loop this chat must be copy-pasted into e-mail
- Actually makes it harder for the uploader to take action

Example

Example of the comments

2020-12-11 at 05:18:53 UTC

matthew.g.wysocki@gmail.com

Journal page number should be 054912, not 054192.

2020-12-11 at 05:11:51 UTC

matthew.g.wysocki@gmail.com

Why is this table separate from Table 4 (which is just different centralities?)

2020-12-11 at 05:11:16 UTC

matthew.g.wysocki@gmail.com

The 5-6 and 6-7 GeV/c bins are missing (which are 90% CLUL)
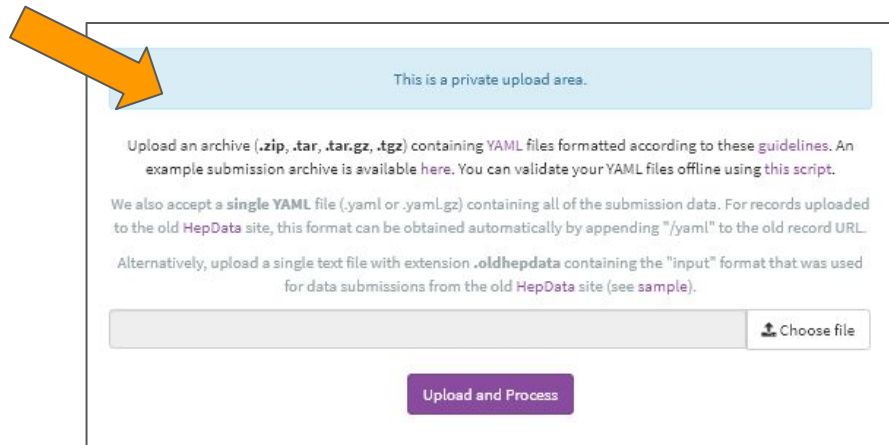
2020-12-11 at 05:10:29 UTC

matthew.g.wysocki@gmail.com

Again, the stat column is actually statistical and type A systematics combined.

Chat vs mail

- PPG doing the upload vs coordinator doing the upload
- Chat allows for interactive discussion
- We could just also switch to direct communication between the reviewer and the designated member of the PPG
- Cost/benefit of the PPG doing the upload - what is the extra workload on the PPG? Considering they need to use the "sandbox" under any scenario, it is not much. The upload itself is trivial
- Opinions?

This is a private upload area.

Upload an archive (**.zip**, **.tar**, **.tar.gz**, **.tgz**) containing YAML files formatted according to these guidelines. An example submission archive is available here. You can validate your YAML files offline using this script.

We also accept a **single YAML** file (.yaml or .yaml.gz) containing all of the submission data. For records uploaded to the old HepData site, this format can be obtained automatically by appending "/yaml" to the old record URL.

Alternatively, upload a single text file with extension **.oldhepdata** containing the "input" format that was used for data submissions from the old HepData site (see sample).

⬆ Choose file

Upload and Process