The PHENIX DAP update

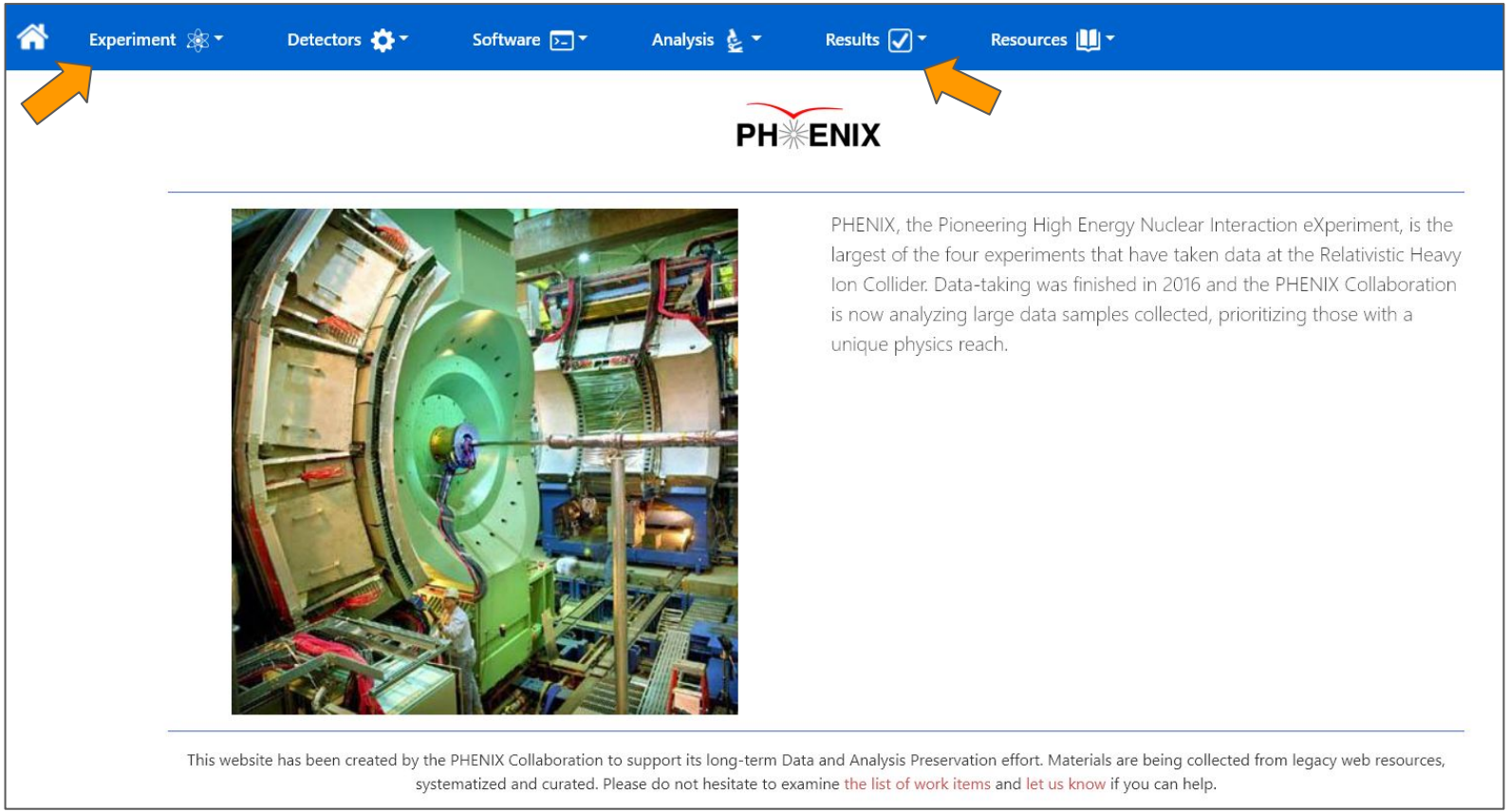
- Presentation at “Future Trends in Nuclear Physics Computing” (9/30/2020)
  - https://doi.org/10.5281/zenodo.4059879
- Please see previous slides for more detail on development in the past month
  - https://github.com/PhenixCollaboration/documentation/blob/master/assets/dap/PHENIX_DAP_20200917.pdf
- Website
- Open Data
- Zenodo
- HEPData/GitHub
- Analysis notes
- PISA
- REANA (?)

Maxim Potekhin
(BNL, NPPS)
10/01/2020

The website

- Site is now live as the official PHENIX resource: https://www.phenix.bnl.gov/
- A few layout problems fixed (resizing problems), menu consolidated
- Created a nice looking “404” page
- “Catch-all” papers.html page to keep this reference live
- The “collaboration” item in the top menu merged into the “Experiment” to save real estate - the top bar was becoming too wide for laptops
- Routine updates of pages (TOF-E etc)
- Conference publications are now links to Zenodo pages (for each conference)

The updated site (deployed)



Open Data

- Gabor and his students prepared Ntuples as material for analysis tutorial, outreach and educational purposes
- I ran simple tests, good stuff, helpful
- Version 2 has been created (smaller number of variables, and a companion note to the Ntuples)
- We can host it on the Open Data Portal (talking to admins now) or on Zenodo
- Well aligned with best community practices

Theses on Zenodo

- Stacyann successfully uploaded a thesis with a good selection of keywords
- The list of keywords has been updated
- The procedure is to add each uploaded thesis to the “documentation” repo to keep track of uploads and also to keep a custodial copy
- More volunteers?

HEPData packages on GitHub

- The HEPData policy in place - good news, lots of work ahead
- We manage submissions on GitHub
- Important for development, testing and preservation of source data
- Instructions will be sent soon to the collaboration mailing list

Redux of a thread in the previous meeting

- Figures and tables (pre-HEPData) - text files are currently stored on the legacy web server
  - Many external links broken for unrelated reasons e.g. missing content
- Would it make sense to ask PPGs to use GitHub to store TXT files? This could make the transition smoother since we already handle HEPData submission packages on GitHub.
  - Even the legacy page could point to GitHub
- A well defined folder structure will make this transparent (e.g. ppg0123)

A “detailed analysis note”

- Write-ups recently added are great (see the analysis section of the site)
- Can a working group be asked to produce a clean and functional step-by-step analysis description?

PISA tutorial

- The legacy tutorial is broken (still under investigation)
- However if one copies materials from the PHENIX repo (e.g. for Run 15) simple examples work (tested)
- So this is the plan how to re-create PISA tutorials, soon to be added to the site

Structured analysis description

- Researchers are learning by example and by borrowing working code
- But in general, each analysis has an “ad hoc” structure
- A more formal description of the workflow would go a long way

REANA (potential cooperation in the community)

- http://reanahub.io/
- “Reproducible research data analysis platform”
- Publicized at CHEP19
  - https://indico.cern.ch/event/773049/contributions/3476160/
  - https://indico.cern.ch/event/773049/contributions/3474811/
  - https://indico.cern.ch/event/773049/contributions/3474839/
  - https://indico.cern.ch/event/773049/contributions/3476165/
- Case in point - structured description of analysis workflows

REANA cont’d

- Cons - a fairly steep learning curve, clearly insufficient resources in PHENIX at this point
- Pros -
  - State of the Art
  - Integration of containers i.e. complete preservation of the analysis environment
  - Highly structured description of analyses is the optimal solution for DAP
  - ...but also increases sustainability of analysis in the medium term - templates can be shared and reused
  - Potential for collaboration with other experiments (e.g. engage with EIC)
- Will take a look at feasibility. How many more PhD theses do we expect?
- There may be more momentum in the NP community in the next few years
- CERN experiments (including ALICE) are already using REANA