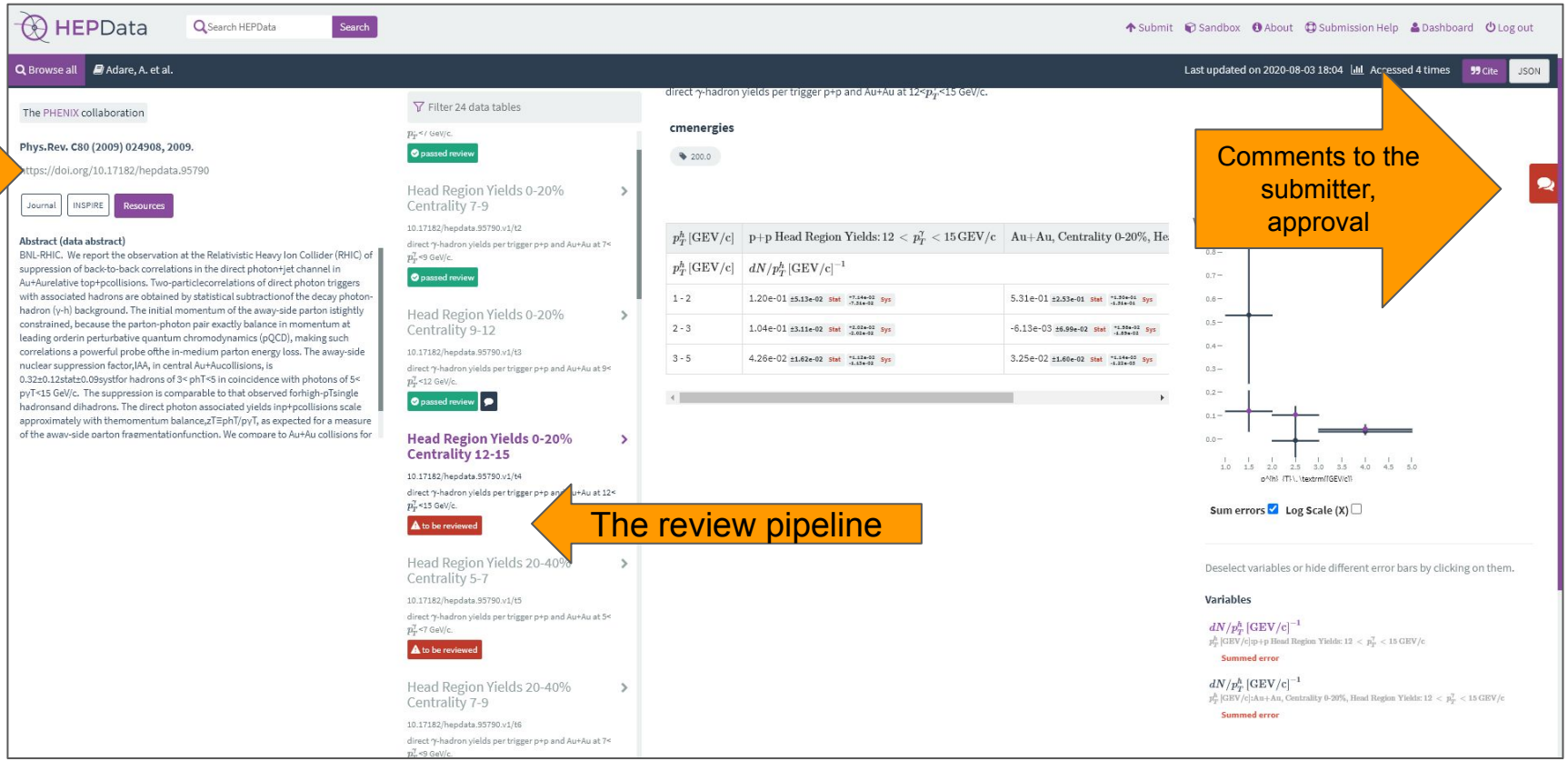
The PHENIX DAP update

Maxim Potekhin
(BNL, NPPS)
08/13/2020

Overview

- HEPData:
  - Thanks to Christine for material preparation
  - Ongoing submission of material previously assembled in the “documentation” repo on GitHub
  - The process is fairly straightforward, tests are possible in the “sandbox”
  - Added a catalog of the PHENIX HEPData entries to the DAP website
- The DAP website:
  - Bug fixes + additional utility code + cleanup
  - Content additions and management (refactoring+cleanup)
  - Working on the components of the Analysis section
  - “Help needed” page added, with a catalog of work items
  - Gabor’s material is in the pipeline
  - Thanks to everyone for recent additions
- A note on GitHub storage policies

The HEPData submission process



HEPData: an example of a finalized submission

Photon-Hadron Jet Correlations in p+p and Au+Au Collisions at s**(1/2) = 200-GeV

The PHENIX collaboration Adare, A. ; Afanasiev, S. ; Aidala, C. ; *et al.*

Phys.Rev. C80 (2009) 024908, 2009.

Inspire Record 815824 DOI 10.17182/hepdata.95790

We report the observation at the Relativistic Heavy Ion Collider (RHIC) of suppression of back-to-back correlations in the direct photor
subtraction of the decay photon-hadron background. The initial momentum of the away-side parton is tightly constrained, because th

24 data tables match query

Head Region Yields 0-20% Centrality 5-7	direct γ-hadron yields per trigger p+p and Au+Au at 5<
Head Region Yields 0-20% Centrality 7-9	direct γ-hadron yields per trigger p+p and Au+Au at 7<
Head Region Yields 0-20% Centrality 9-12	direct γ-hadron yields per trigger p+p and Au+Au at 9<

More...

The HEPData catalog on the DAP website

PPG references for new uploads

PH ENIX | Experiment | Detectors | Software | Analysis | Resources | About

HEPData

HEPData is an open-access repository for scattering data from experimental particle physics which includes data points from several thousand publications.

The PHENIX Collaboration is using this platform as one of the components of its Data and Analysis Preservation (DAP) effort, and is adding material to its PHENIX on HEPData. Members of the PHENIX Collaboration interested in creating materials suitable for submission to HEPData are encouraged to contact the DAP Team.

There is a helpful write-up about preparing data for upload:

- DOI 10.5281/zenodo.3974601 How to make HEPData input (C. Nattrass)

Listed below are PHENIX entries uploaded to the HEPData website.

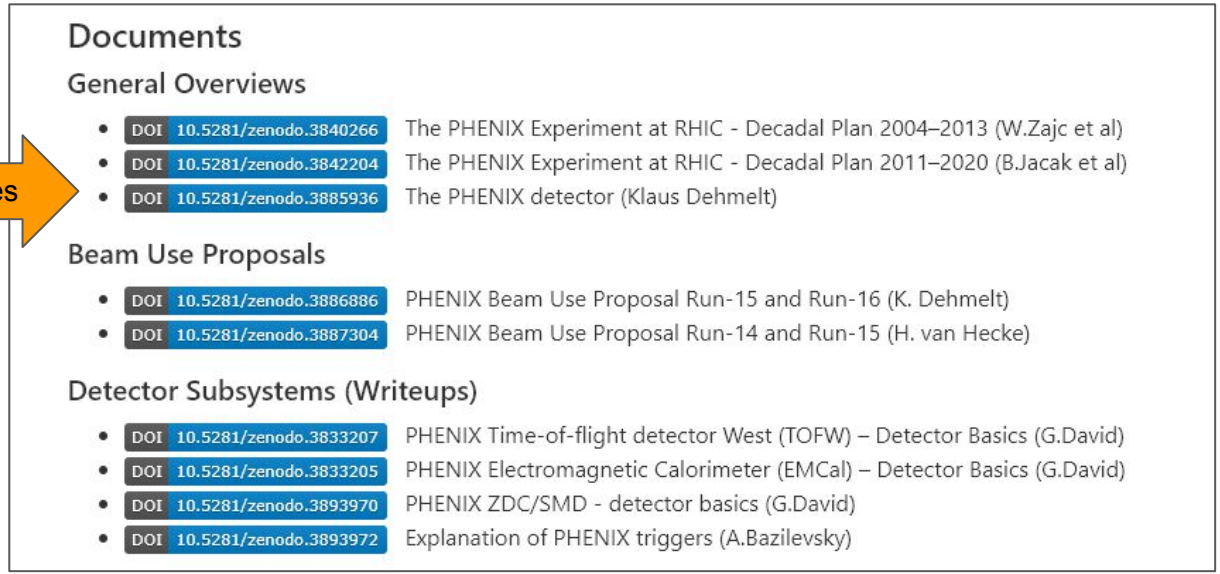
Publication Materials

- HEPData 95883 Measurement of charged pion double spin asymmetries at midrapidity in longitudinally polarized p+p collisions at $\sqrt{s}$=510GeV : **arxiv 2004.02681 (ppg227, inspire 1789851)**
- HEPData 95877 Medium modification of jet fragmentation in Au + Au collisions at $\sqrt{s_{NN}}$=200GeV measured in direct photon-hadron correlations : **Phys.Rev.Lett. 111 (2013) 032301, 2013 (ppg113, inspire 1207323)**
- HEPData 95866 High p_T direct photon and π^0 triggered azimuthal jet correlations and measurement of k_T for isolated direct photons in p+p collisions at $\sqrt{s}$=200GeV : **Phys.Rev. D82 (2010) 072001, 2010 (ppg095, inspire 857187)**
- HEPData 95817 High-p_T π^0 Production with Respect to the Reaction Plane in Au + Au Collisions at $\sqrt{s_{NN}}$=200GeV : **Phys.Rev. C80 (2009) 054907, 2009 (ppg092, inspire 816486)**
- HEPData 95815 Quantitative Constraints on the Opacity of Hot Partonic Matter from Semi-Inclusive Single High Transverse Momentum Pion Suppression in Au+Au collisions at $\sqrt{s_{NN}}$=200GeV : **Phys.Rev. C77 (2008) 064907, 2008 (ppg079, inspire 777211)**
- HEPData 95790 Photon-Hadron Jet Correlations in p+p and Au+Au Collisions at $\sqrt{s}$=200GeV : **Phys.Rev. C80 (2009) 024908, 2009 (ppg090, inspire 815824)**

The DAP site: development + cleanup

- A few bugs cropped up during the development, fixed these
- Switched to the new interfaces throughout the site
- Removed extraneous pages with little material on them
- Fixed server error 429 coming from Zenodo (now caching badges)

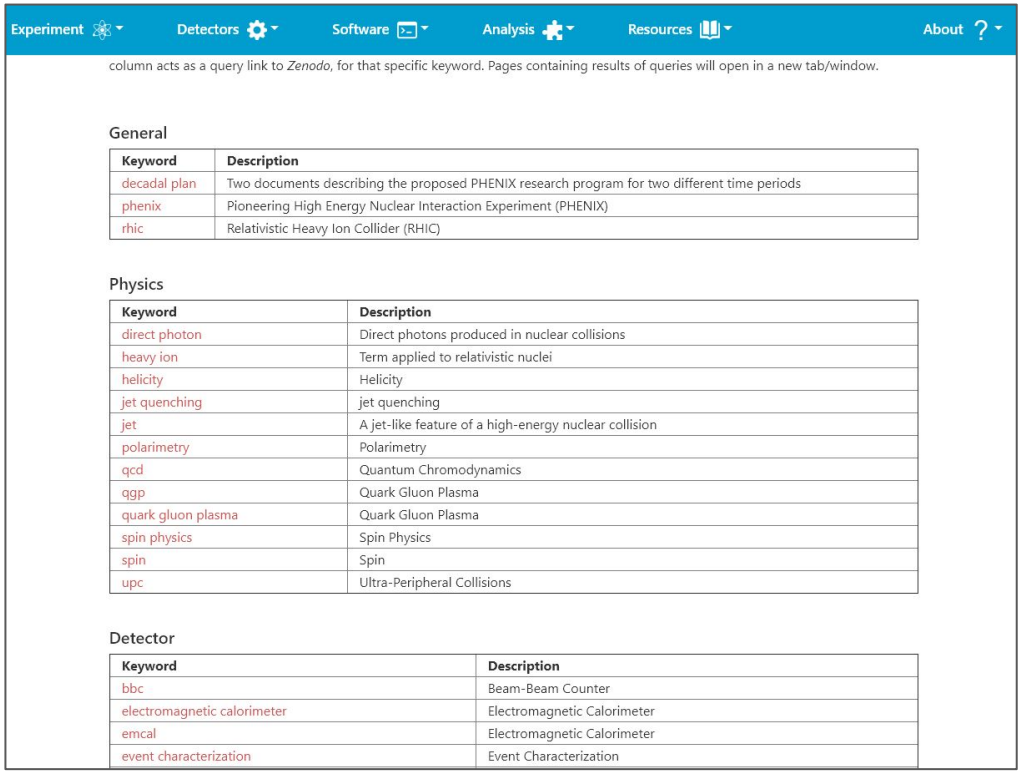
Zenodo DOI badges



The DAP site: refactoring

- Continued to normalize data i.e. reference only central registries as opposed to the front matter (YAML) or content on pages, removed obsolete macros
  - e.g. moved all of the run data to runs.yml from the Front Matter YAML individual pages, removed hard links to images in favor of central registry
  - Removed old code, added a few internal navigation macros
- The keywords dictionary and page
  - The keywords page was becoming quite long and confusing
  - Created category-specific tables (detector, physics, runs etc)
  - Added many relevant keywords for physics, detector etc (including "J/Ψ")
- Merged the “DAP” and “site” pages to make space for “Help needed”
  - Ron’s suggestion, hopefully will make it easier to solicit help

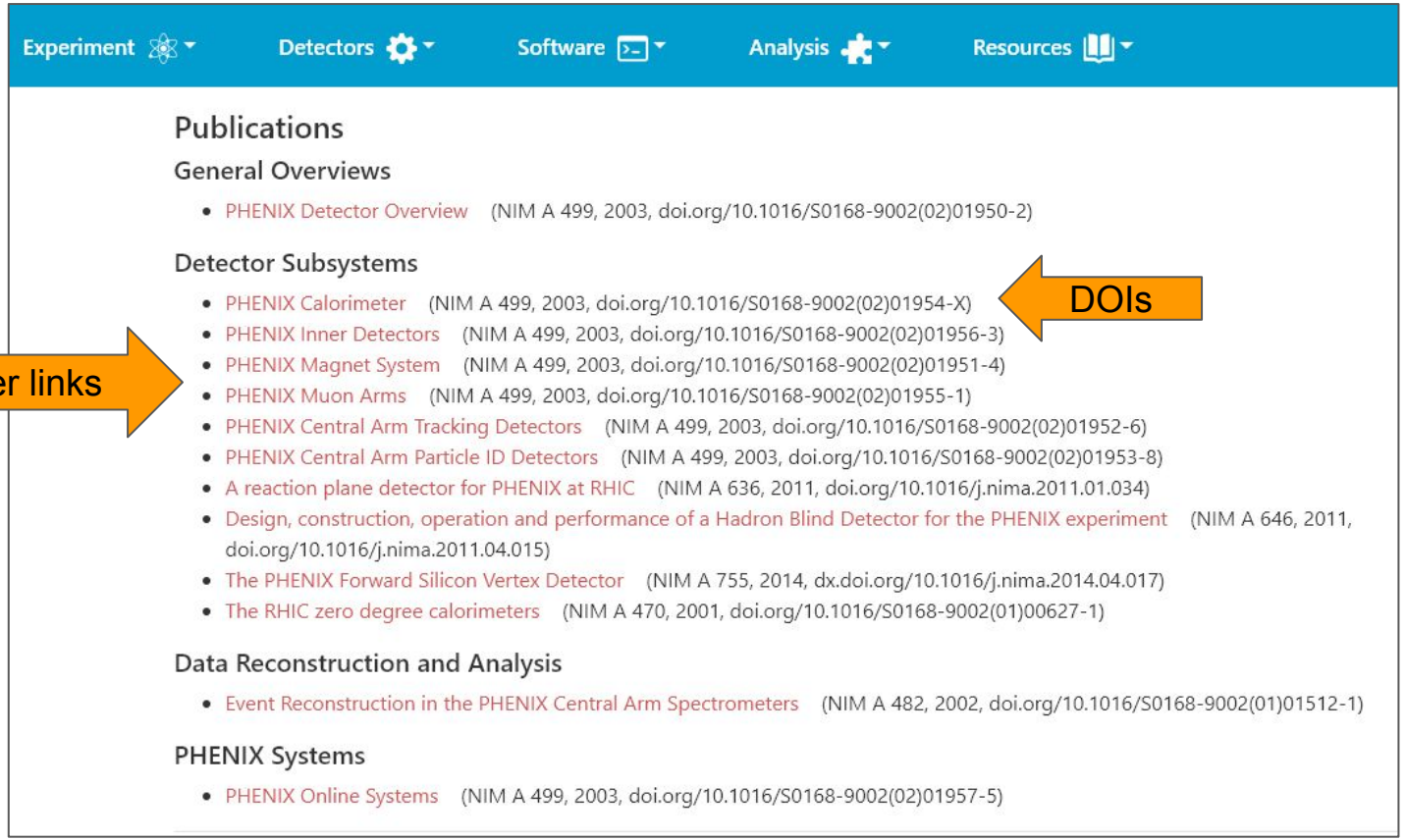
The keywords page updated



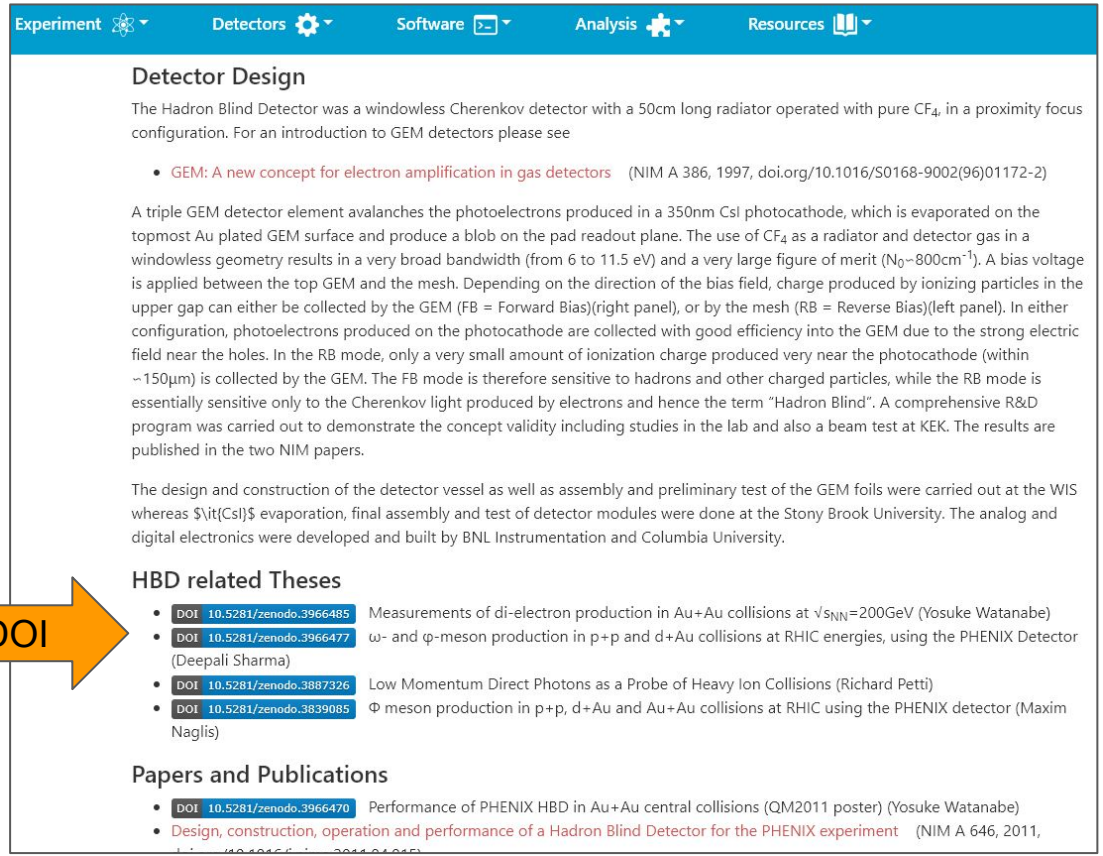
Papers and theses

- As agreed a while ago, removed PDFs of the **NIM papers** from the public area of the site - replaced with proper DOIs
- Migrated to our separate “documentation” repository as encrypted files
  - The standard Linux “mcrypt” used
  - The “usual” DAP password (ask me if you forgot)
  - https://github.com/PhenixCollaboration/documentation/tree/master/assets/papers
- Removed links to PDFs hosted on the legacy phenix.bnl.gov site (theses etc)
- Continued uploading theses (as far as is possible with the current server status) and adding relevant links to pages
  - Total of 15 uploaded
  - e.g. ZDC, HBD etc
  - AFAIK there is no dedicated thesis folder on the legacy server which makes this hard

NIM papers - migration (PDFs removed)



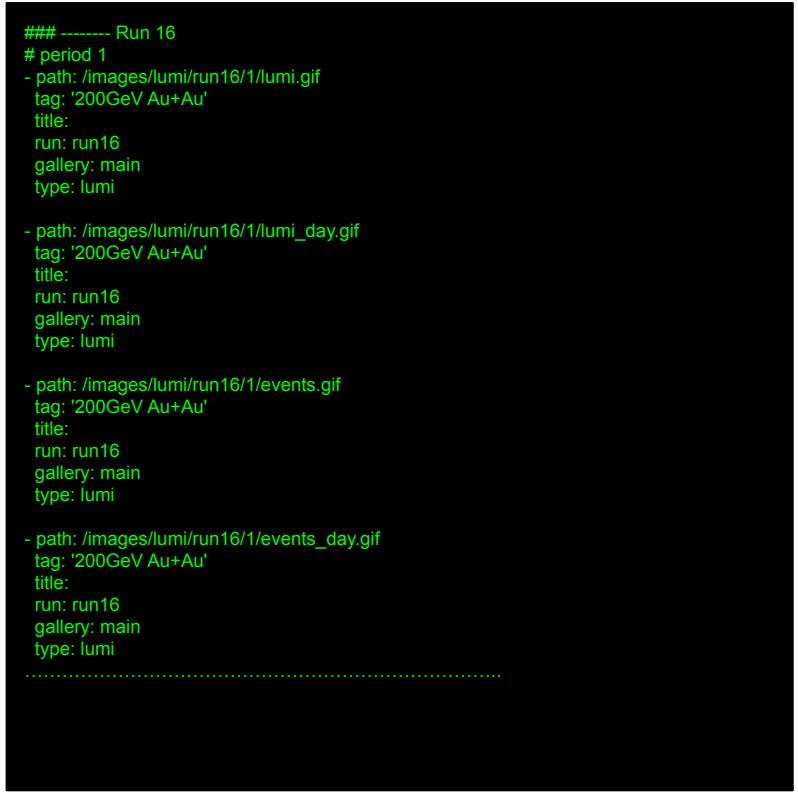
The revamped HBD page



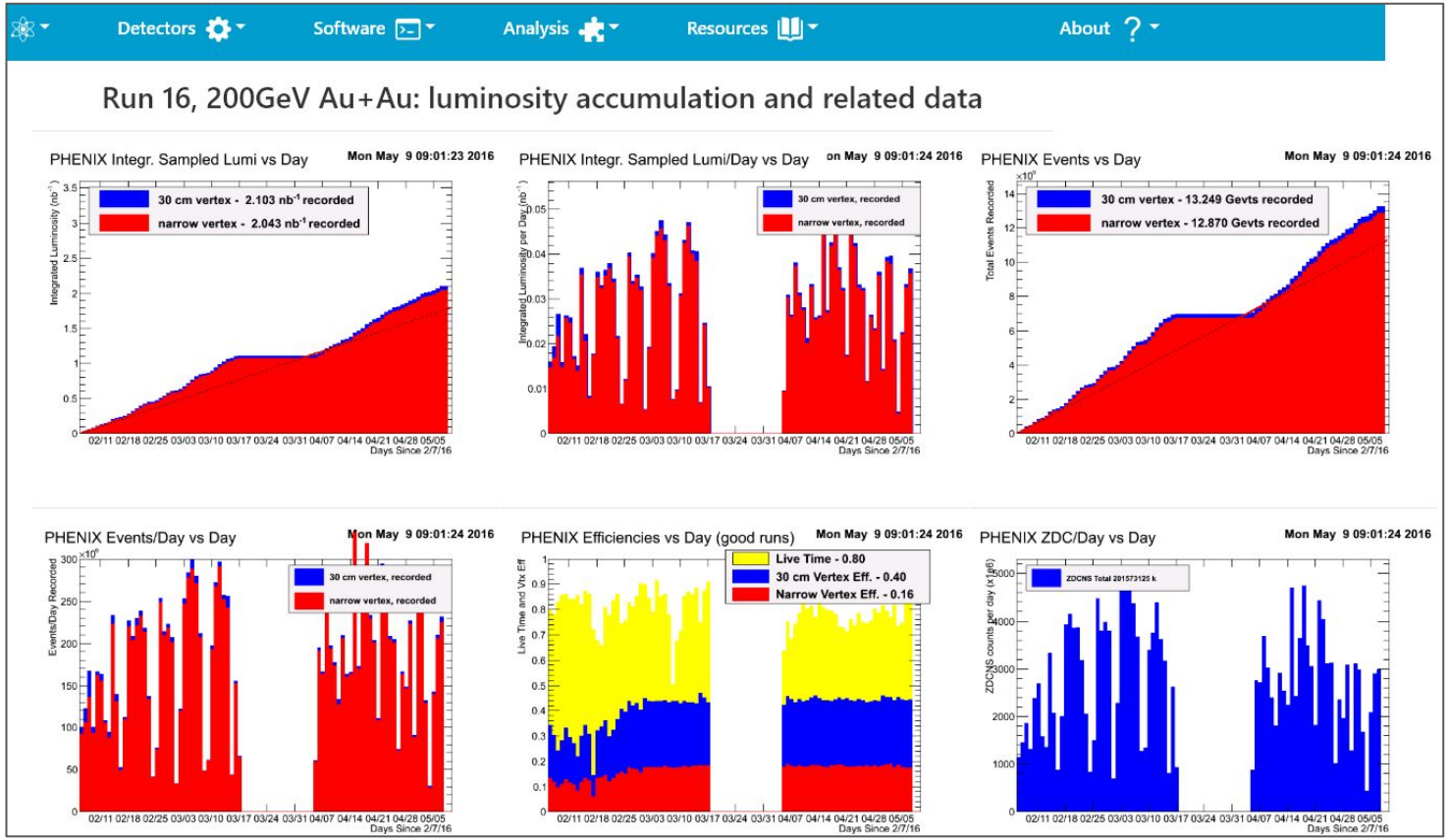
Links: all DOI

The DAP site content: lumi plots

- Resumed migration of the lumi content
  - Take plots from the old run pages
  - Add plot metadata to the gallery registry
  - Not challenging but still time consuming (lots of plots)
- Addition of luminosity data with custom periods
  - Had to augment the structure of the run registry and some logic to account for multiple periods not mapped to official subruns (i.e. AuAu has two separate periods in run 16)
  - 6 individual periods
- Added plots for runs 09 through 16
  - Total of ~200 plots, size is not an issue



Lumi section updates (sample)

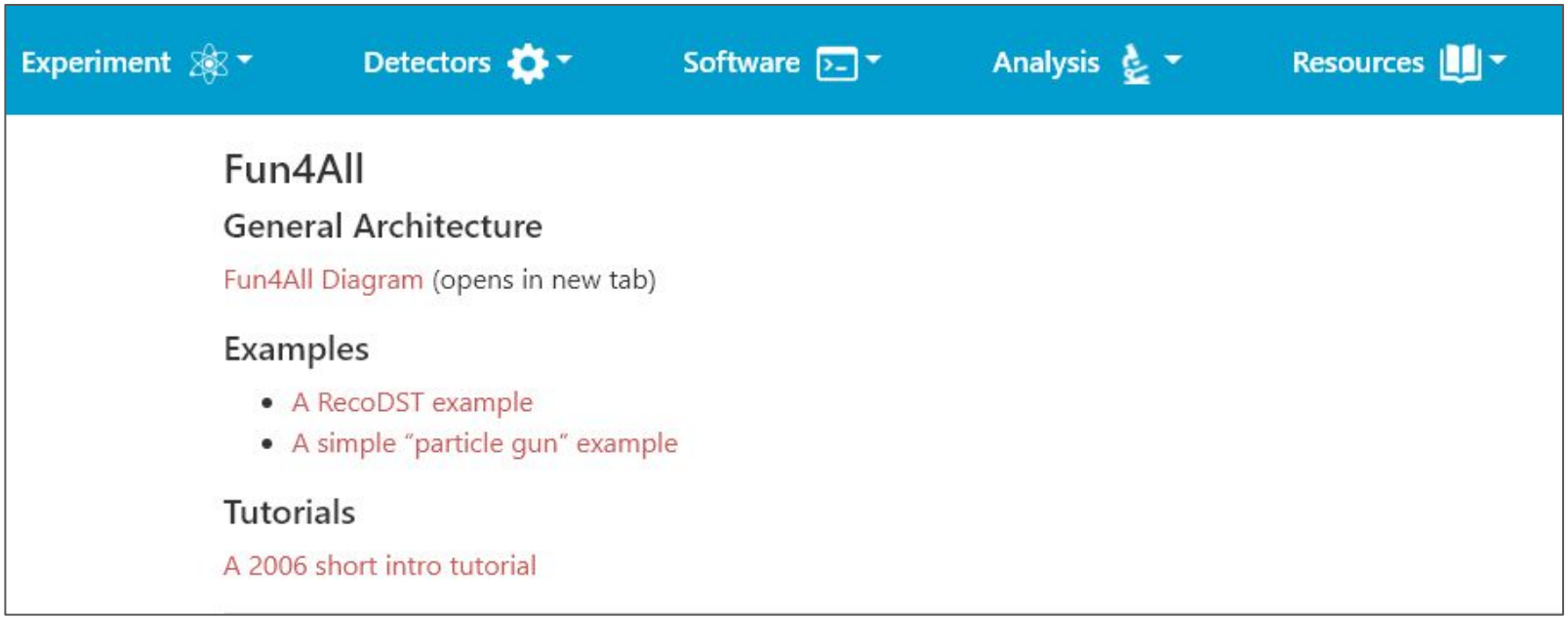


Luminosity Data etc

- Does it make sense to extract, index and preserve actual luminosity data currently in the database?
- Any other suggestions about what other plots or data need to be migrated from the old hosts to the DAP site?

The DAP site content: software section

- Added some material from the Wiki to the Overview
  - https://phenixcollaboration.github.io/web/software/overview.html
- Improved the Fun4All page and added a simplified particle generator example
  - https://phenixcollaboration.github.io/web/software/fun4all.html



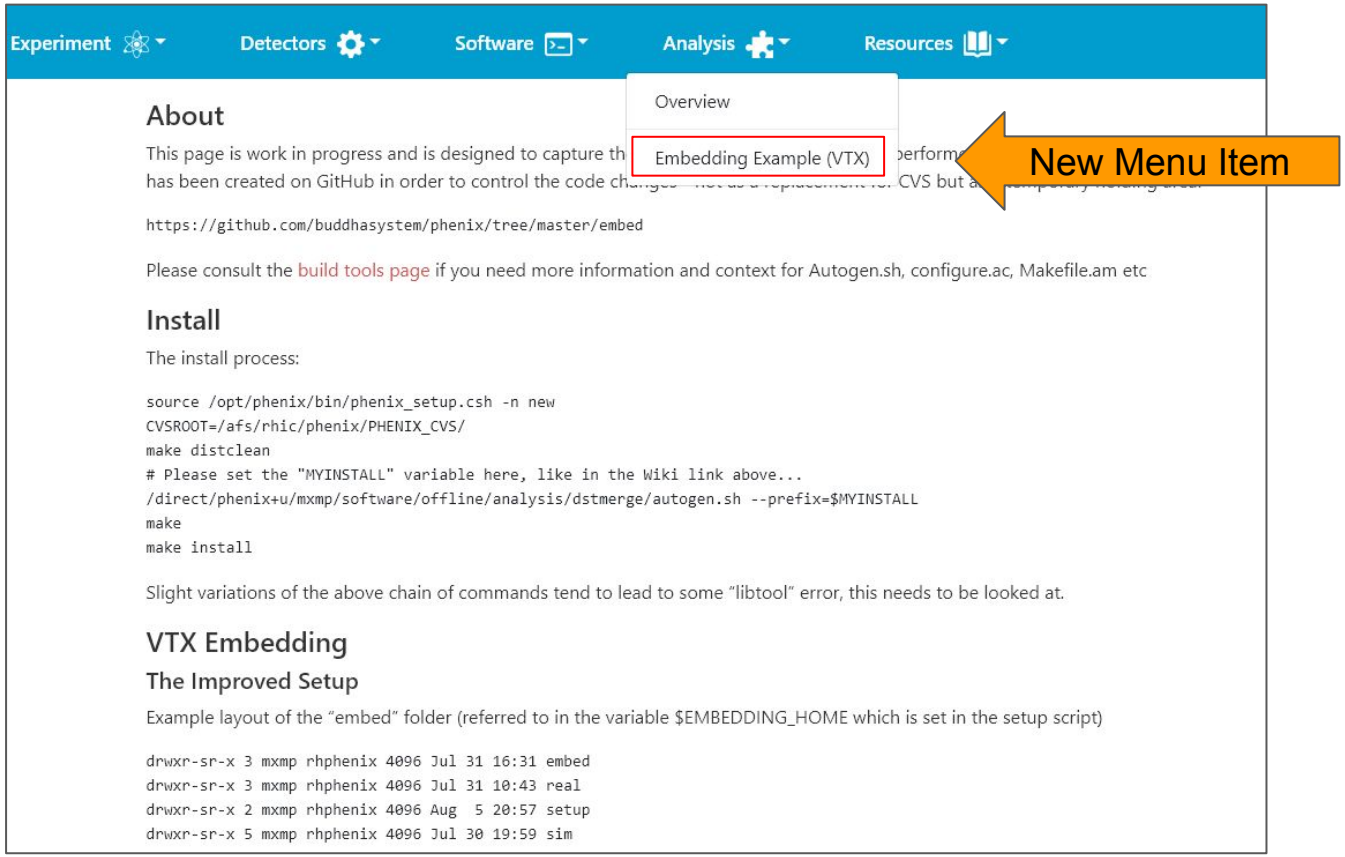
Particle generator example also available on GitHub

```
//
// A few additional comments/annotations were added.
//
//
void
phparticlegen_example(const int nevents = 10, int pid=11, const char *outputname = "phparticlegen_example.root")
{
  gSystem->Load("libPythia6.so");
  gSystem->Load("libfun4all.so");       // framework + reco modules
  gSystem->Load("libpdbcalBase.so");
  gSystem->Load("libPHParticleGen.so");
  gSystem->Load("libPHPythia.so");
  gSystem->ListLibraries();

  recoConsts *rc = recoConsts::instance(); // fun4all/PHOOL singleton class
  rc->set_IntFlag("RUNNUMBER",0);

  //  Instantiate a Fun4All Server
  Fun4AllServer *se = Fun4AllServer::instance();

  // A custom generator can be added here
  PHParticleGen* ss = new PHParticleGen();

  // Can optionally set the seed, otherwise defaults to /dev/random
  // ss->SetSeed(1999);

  //*** Set Particle Parameters. Overrides config file.

  // Particle ID
  ss->GetGenerator()->SetParameter("pid", pid); // e+

  // Momentum/Transverse Momentum/Energy
  ss->GetGenerator()->SetParameter("momflag",TSingleParticleGenerator::FLAT_PT);
  ss->GetGenerator()->SetParameter("pmin",0.5);
  ss->GetGenerator()->SetParameter("pmax",10.);

```

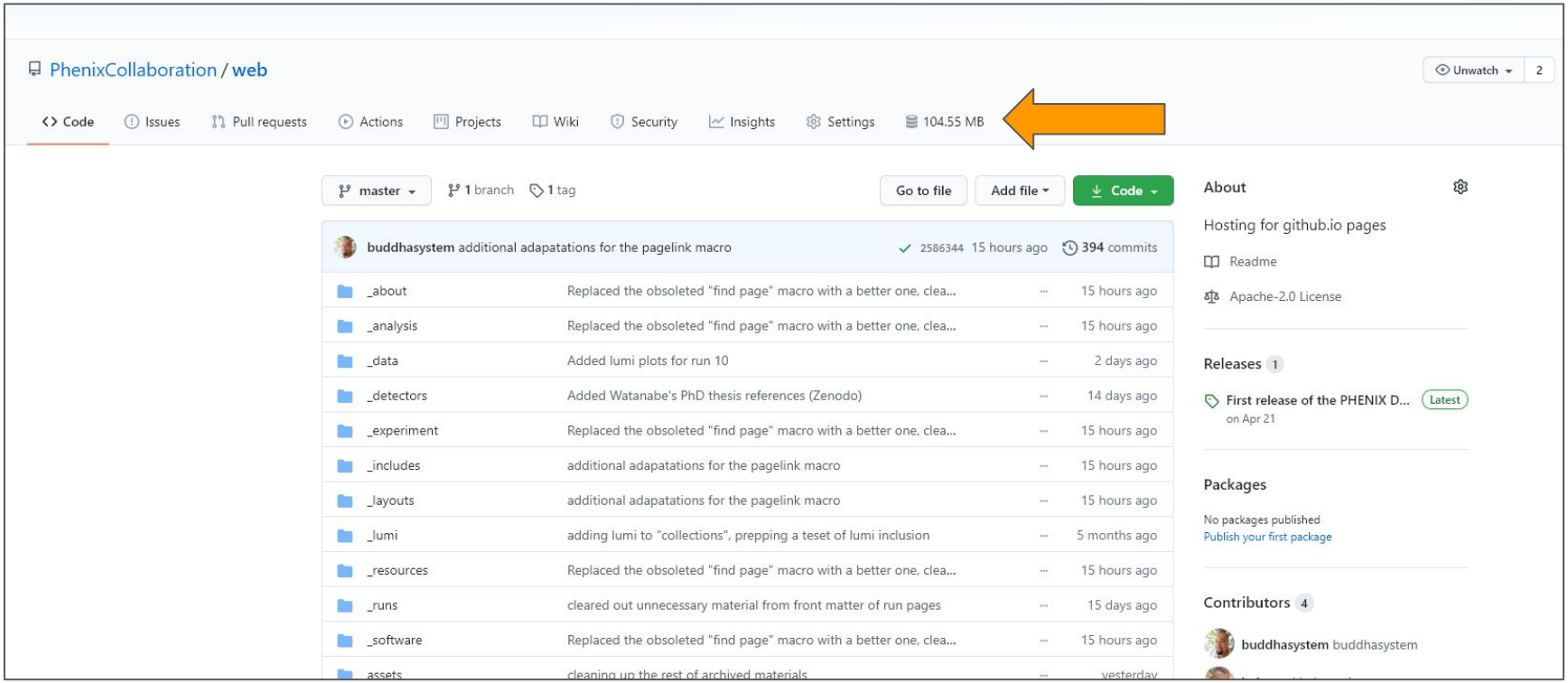
The DAP site content: analysis

- In 2019 a survey of embedding procedures in two separate analyses was performed, and the notes were placed in the Wiki
- Now, started migration of the notes to the DAP site
  - Working to improve the code along the way - still finding extraneous dependencies
  - A “scratchpad” repo on GitHub to keep the snapshot and corrections to it
  - Embedding done by Takashi - the analysis seems like a good candidate to be refined and placed on the DAP site, also consider Takashi’s availability to consult
- Potential to create functioning tutorials
  - Created a “software” section in the **documentation** repo (not to be confused with the “web” repository where we keep the site), with “examples” section
  - Added a simple fun4all example - the “phparticlegen” macro used in some analyses
  - This is aligned with our previous discussions

The embedding note prototype on the site



Size of the DAP website repo



Notes on repo sizes and policies

- Depending on the client, GitHub caps *generic* repositories at 75 to 100GB
- For *GitHub pages* (our current hosting solution) there are separate caps:
  - 1GB for the repo from which the site is built (currently 100MB)
    - I freed ~20MB in the past few days by moving materials to Zenodo
  - 1GB for the compiled site content (currently 50MB)
  - 100GB/month traffic - doesn't see to be an issue
- Caveats
  - The size of a local clone of the repo will be different since the folders also contain the git "objects" i.e. the database which keeps previous commits. Can be fixed, an advanced subject.
- Going forward
  - Leverage Zenodo, GitHub and potentially OpenData to offload anything "big" from the site, avoid adding to the site repo in the first place to keep the git database lean
  - **We'll be fine** with regards to the GitHub limits if we are careful
  - Small images/diagrams should be fine, but no large PDFs etc

Misc

- Missing parts of the PHENIX (internal) website are still a bit of a pain
  - Analysis notes pages are not fully functional, no easy access to theses etc
- Support/updates of *legacy* services (phenix.bnl.gov) may become a serious issue a few years down the road
  - PHP ecosystem is subject to disruptive version changes, to be built vis a vis evolving OS's
  - More motivation to migrate to the new site as much and as soon as possible
  - Can we survey/identify the parts that need to be migrated?
- Survey of analysis notes
  - Some better than others, but many don't comply with the template
  - ...some are missing reference to the analysis code
  - Exact locations of the data components are often missing
  - Software: no clear how-to instructions in all notes
  - In summary, analysis notes will not be of much use for DAP - must rely on direct communication with the researcher

Summary and plans

- The first “new” batch of HEPData submissions is being uploaded, the process is working well
  - Presentation for the EC is ready (Christine and Maxim)
- Progress is being made with the content of the DAP site and its functionality
  - All PDF content on Zenodo
  - NIM papers migrated to a separate repository, with proper DOIs added to the site
  - Will only upload moderate size images and diagrams directly to the site, all else to Zenodo etc
- Most lumi plots migrated to the DAP site
- The embedding “use case” (Takashi) is promising and work shall continue to develop it and leverage to document the PISA and other components