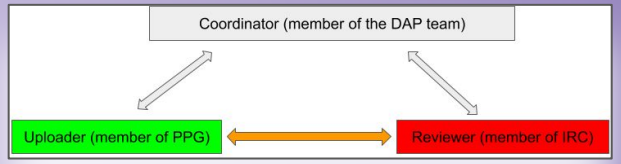
Data & Analysis Preservation: current work items



Maxim Potekhin
Nuclear and Particle Physics Software Group

BROOKHAVEN NATIONAL LABORATORY

PHENIX

PHENIX DAP Meeting
03/11/2021

Overview

- Zenodo/Website
- DB
- Open Data
- HEPData
- Analysis notes decision
- ACAT 2021 abstract: DAP in PHENIX

Zenodo/Website

- Ongoing work by Gabor on Zenodo uploads/keyword management
- Web: we operate two sites, one for development and another in production (@BNL)
- *The Development website* (@GitHub) - https://phenixcollaboration.github.io/web/
  - Caveat: functionality may be erratic when accessed via the BNL VPN, something to note but not of much concern since majority of access is via direct access
  - Likely culprit is occasional blocking of standard JS and CSS
  - The production website **phenix.bnl.gov** is not affected
- Concern - haven't added new materials to the website in a while

DB

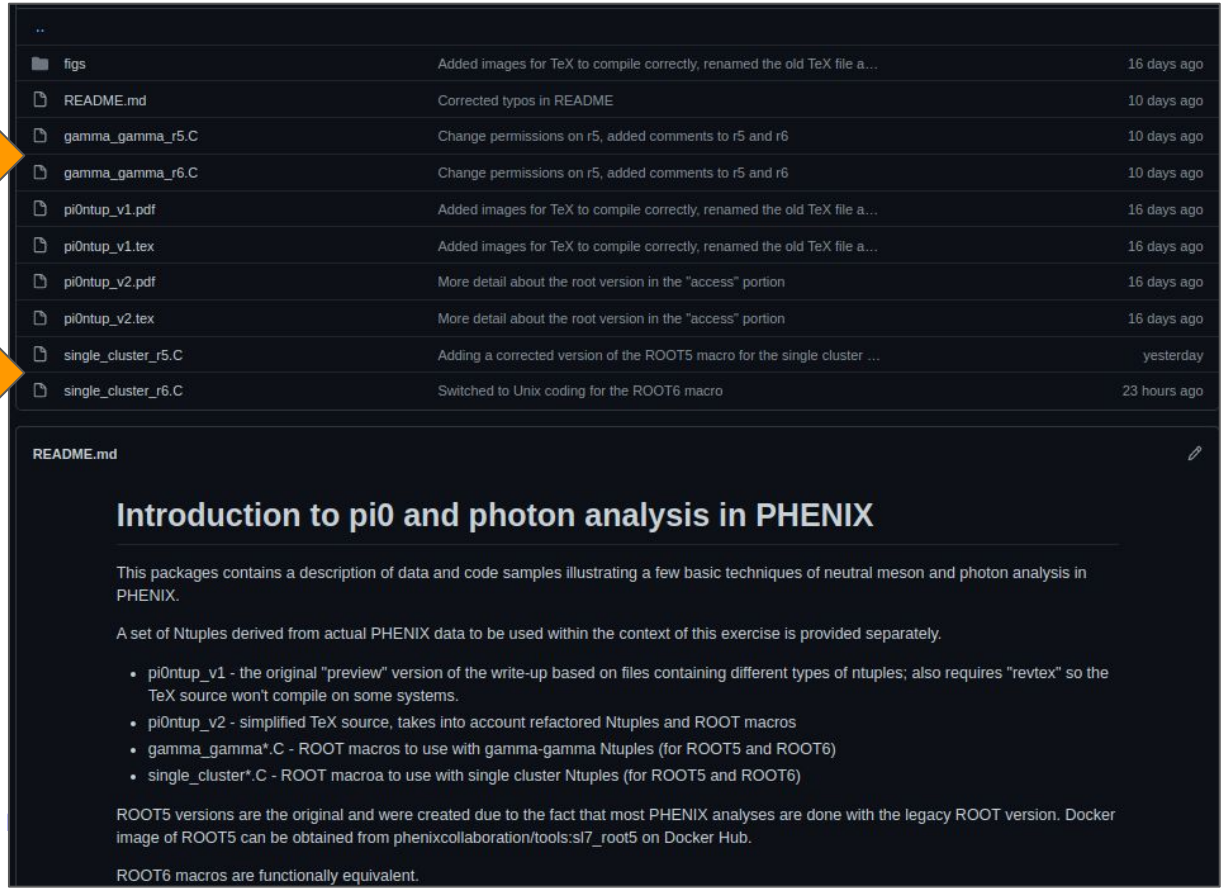
- Handover (?) of DB cron jobs on a legacy server - Irina's request to Chris
- Scripts written in Perl, and right now we don't know what they do
- Will need a survey to understand functionality and importance
- Up till now ran from an individual user's directory (by Irina)
- Need to concentrate on critical functionality which can't be migrated (i.e. older analysis notes) and speed up migration elsewhere
- Could be easier/more reasonable to migrate ~1700 analysis notes than debug and maintain legacy scripts and stored procedures
- Availability of effort is a problem
  - On that note, are we getting a postdoc with a DAP role?
  - Any prospects of support from the Lab?

Open Data - the “π^0 and ɣ analysis” entry

- Realized that having only ROOT5-compatible macros in the package was a little awkward even with the included legacy image
- Created and tested parallel ROOT6 macros, compared every plot with ROOT5
  - ...by using our ROOT5/SL7 Docker image
- Will push submission package to CERN this week
- Can we find volunteers/create new entries?
  - Try to canvass the conveners?
  - The current model of NTuple+ROOT is a low-hanging fruit, fits into DAP methodology
  - Will help create useful material for the School
- Pretty much done!

Our Open Data entry on GitHub (updated)

https://github.com/PhenixCollaboration/opendata/tree/master/pi0_photon_analysis



..

figs	Added images for TeX to compile correctly, renamed the old TeX file a…	16 days ago
README.md	Corrected typos in README	10 days ago
gamma_gamma_r5.C	Change permissions on r5, added comments to r5 and r6	10 days ago
gamma_gamma_r6.C	Change permissions on r5, added comments to r5 and r6	10 days ago
pi0ntup_v1.pdf	Added images for TeX to compile correctly, renamed the old TeX file a…	16 days ago
pi0ntup_v1.tex	Added images for TeX to compile correctly, renamed the old TeX file a…	16 days ago
pi0ntup_v2.pdf	More detail about the root version in the "access" portion	16 days ago
pi0ntup_v2.tex	More detail about the root version in the "access" portion	16 days ago
single_cluster_r5.C	Adding a corrected version of the ROOT5 macro for the single cluster …	yesterday
single_cluster_r6.C	Switched to Unix coding for the ROOT6 macro	23 hours ago

README.md

Introduction to pi0 and photon analysis in PHENIX

This packages contains a description of data and code samples illustrating a few basic techniques of neutral meson and photon analysis in PHENIX.

A set of Ntuples derived from actual PHENIX data to be used within the context of this exercise is provided separately.

- pi0ntup_v1 - the original "preview" version of the write-up based on files containing different types of ntuples; also requires "revtex" so the TeX source won't compile on some systems.
- pi0ntup_v2 - simplified TeX source, takes into account refactored Ntuples and ROOT macros
- gamma_gamma*.C - ROOT macros to use with gamma-gamma Ntuples (for ROOT5 and ROOT6)
- single_cluster*.C - ROOT macroa to use with single cluster Ntuples (for ROOT5 and ROOT6)

ROOT5 versions are the original and were created due to the fact that most PHENIX analyses are done with the legacy ROOT version. Docker image of ROOT5 can be obtained from phenixcollaboration/tools:sl7_root5 on Docker Hub.

ROOT6 macros are functionally equivalent.

HEPData

- A new script - to aid in formatting the errors - has been developed using the original script developed in STAR
- Sharing OK'd by Rongrong Ma, will be credited
- Will test, add to the repository, make improvements as necessary

Analysis notes management options: the shortlist

- GitHub (cloud) vs Gitea (BNL)
- GitHub is considered because it’s ubiquitous
- Gitea is in the picture due to expected commitment on the part of BNL for long-term support

Analysis notes options: GitHub

- GitHub - can use a private repository to host analysis notes
  - Accessible to users on a managed list
  - GitHub tags can be used for indexing (like keywords)
  - Concerns about longevity of this resource (but should we be concerned?)
  - Leading provider, highly rated, widely used
  - ...if it goes out of business we'll be in the company of world leading enterprises migrating to the next platform
  - Will it be around in 10 years? My guess would be yes.
  - NB. Data won't be lost in any case - we'll have local repo clones, long term backup solutions such as HPSS and other BNL storage etc

Today 52 percent of Fortune 50 companies use GitHub's Enterprise business tier, which costs $21 per user per month. Altogether, GitHub has more 23.1 million users in 200 countries and 1.5 million organizations.

Analysis notes options: Gitea

- Gitea at BNL
  - Based on the same underlying repo technology (git)
  - A different product and not a cloud service like GitHub
  - Another layer of access control (controlled by BNL)
    - This can be good or bad
  - No free hosting of web site(s), some other differences
  - We can *hope* for long-term commitment of our organization to this service

- My personal vote is for GitHub

ACAT(Advanced Computing and Analysis Techniques) 2021

“The 20th edition of ACAT will bring together experts to explore and confront the boundaries of computing, automated data analysis, and theoretical calculation technologies, in particle and nuclear physics, astronomy and astrophysics, cosmology, accelerator science and beyond.”

- A prestigious international conference with a history of BNL participation
- Will result in a high quality publication
- A good way to highlight our participation in the global DAP community, emphasize our leverage of state-of-the art technologies
- Draft abstract created - comments?