

PHENIX DAP Update: The site, Zenodo Communities and GitHub Integration

Maxim Potekhin
(BNL, NPPS)
04/23/2020

The DAP website

- Deepali made a contribution to the HBD page
 - Thank you. The pull request process worked fine.
- Improvements in the site mechanics are in the works:
 - Better organization of menus, making it easier to add material - just plug in entries in the menu descriptor instead of tweaking weights in each file
 - Will be ready in a day or two
- A template for the analysis pages - this still requires thought
- Referencing analysis code on the DAP site:
 - See the discussion of **GitHub and Zenodo** in the following slides
 - But also a matter of policy: how much of the code do we publish on the site?
 - cf. the site is public
 - A small fragment will not achieve anything since the goal is reproducibility

Zenodo

- Recap of previous presentations:
 - Low entry cost
 - Metadata support
 - Can store datasets, code or any other digital products
 - Generates and supports official DOIs (i.e. doi.org)
 - Prime candidate for use in PHENIX
- Today's Zenodo items:
 - Communities
 - Communication with the developers
 - GitHub integration


Zenodo Community


- A way to organize material, and to consistently attribute materials to a collaboration/project/experiment - keeping a consistent brand
- An improvement in visibility/discoverability/PR
 - An addition to the already existing metadata query aids in discovery of materials and in practice works better
- Anyone can upload a material to the community which is subject to **curation**
 - The curator gets notified and inspects the submission
 - If accepted, it becomes posted under the community umbrella
 - If rejected, it still remains on Zenodo site but is not officially owned/acknowledged by the community, this is an accordance to the “open access” platform
 - *There is currently one curator per community and there is no easy way to transfer this duty to a different account* (something few people expected) but a fix is on the way according to the lead developer and other team members. Unofficial ETA is late 2020.

Zenodo “PHENIX community” - initial test

- As an initial test (which can be extended) I created a community for the PHENIX Collaboration under the tag “**phenix-collaboration**” similar to our organization name on GitHub
 - Also, “phenix” was already taken by a French bio lab
- As with most other features of Zenodo, the setup was fairly easy

Zenodo Community Example: the PHENIX proto

QUploadCommunities

 potekhin@bnl.gov

The PHENIX Collaboration Community

Recent uploads

Q

April 21, 2020 (v1.0) Software Open Access

PhenixCollaboration/web: First release of the PHENIX DAP site

Maxim Potekhin; Ron Belmont; amolhj

This is the first release of the PHENIX DAP website

Uploaded on April 21, 2020

View

May 1, 2019 (v1) Thesis Open Access

Transverse, Single-Spin Asymmetries for Charged Hadrons and for Muons from Open-Heavy-Flavor Decays in Polarized Proton-Proton and Proton-Nucleus Collisions in PHENIX

Bok, Jeongsu.

Transverse single-spin asymmetry (TSSA) phenomena have gained substantial attention in several decades because they provide valuable information on the spin structure of the nucleon. Production of heavy flavor is dominated by gluon-gluon fusion in the leading order perturbative Quantum Chromodynamic

Uploaded on April 20, 2020

View

May 1, 2019 (v1) Thesis Open Access

Measurements of $\mu\mu$ pairs from cc , $b\bar{b}$ and Drell-Yan in p+p and p+Au collisions at $\sqrt{s_{NN}} = 200$ GeV with PHENIX at RHIC


Leung, Yue Hang.

Dilepton spectra are a classic probe to study ultra-relativistic heavy ion collisions. At $\sqrt{s_{NN}} = 200$ GeV, the dilepton continuum is dominated by correlated pairs from semi-leptonic decays of charm and bottom hadrons and the Drell-Yan process. Measuring the azimuthal correlations of heavy flavor


Uploaded on April 20, 2020

View

More

 New upload

Community



The PHENIX Collaboration Community

The PHENIX Collaboration has initiated a Data and Analysis Preservation (DAP) effort in 2019. Within this scope there are a few areas of activity such as curating the available information, development of a new website to support DAP and to systematize and store available document for the long term. The latter is the main reason the PHENIX Zenodo community has been created.

Curated by:

MaximPotekhin

Curation policy:

Curation is done by members of the PHENIX DAP Task Force and contributors who join individual projects within the scope of DAP.


Created:

April 20, 2020

Harvesting API:

GAI-PMH interface

Want your upload to appear in this community?

 Click the button above to upload straight to

Zenodo issues and next steps

- I believe it was agreed that we need to start uploading and tagging theses
 - Already started doing this as a part of the test
- Having these materials posted under the PHENIX Community umbrella seems like precisely the right solution
- Are we OK with having me (Maxim) as the curator of the community?
 - I'd rather not distribute my own credentials
- Alternatively, the prototype community can be deleted and a new production account created on Zenodo with multiple people having access to it
 - this creates a separate set of (minor) problems TBD e.g. with the GitHub link

Zenodo uploads, going forward

- I need help with uploads of theses and other materials
 - Assigning metadata (keyword) to documents is best done by domain experts
 - Students/Grad Students/Other community members can easily contribute
 - Low-hanging fruit, real results visible immediately
 - Curation makes it easy to accept uploads from anyone - any uploads, of any kind
 - If theses authors do this themselves this is even better
 - Should also benefit from the ORCID reference mechanism
- What other document (in addition to theses) should be considered public?
 - We need to reduce storage requirements for the GitHub repo of the site, and there are some large documents there already which really belong in the long-term document database
 - Of course can also create a separate repo and will probably do so
 - But for example should we just move the decadal plans and detector writeups to Zenodo?

GitHub/Zenodo mechanics


- A snapshot of a GitHub repository can be included in Zenodo organically, and a DOI generated
 - Prepares and preserves tarballs of your releases
 - Makes your code easy to find (using the metadata) and to reference by a unique ID
 - Nice GUI
 - DOI reference to analysis code
- Easy to use
 - I tested this functionality and it was quite simple
 - DOIs take some time $O(10\text{min})$ to propagate to the DOI.org system


Zenodo - GitHub panel - repo selection

The screenshot displays the Zenodo GitHub panel interface. At the top, the Zenodo logo is on the left, followed by a search bar and links for 'Upload' and 'Communities'. The user's email, 'potekhin@bnl.gov', is shown in the top right. Below the header, a breadcrumb trail reads 'Home / Account / GitHub'. On the left, a 'Settings' sidebar lists options: Profile, Change password, Security, Linked accounts, Applications, Shared links, and GitHub (which is selected). The main content area is titled 'GitHub Repositories' and includes a '(updated now)' status and a 'Sync now' button. It features a 'Get started' section with three steps: 1. Flip the switch (with an 'ON' button), 2. Create a release, and 3. Get the badge (showing an example DOI: 10.5281/zenodo.8475). Below this is a 'Repositories' section with a note about third-party access. A list of repositories follows, each with a GitHub icon, the repository name, and an 'OFF' toggle switch:



Repository	Status
BNLNPPS/BNLNPPS.github.io	OFF
BNLNPPS/BirdView	OFF
BNLNPPS/tpc-rs	OFF
DUNE/FNALCore	OFF
DUNE/Sandbox-TDR	OFF
DUNE/SpaceCharge	OFF

Zenodo - GitHub panel - published release







[Upload](#) [Communities](#)


 potekhin@bnl.gov 


[Home](#) / [Account](#) / [GitHub](#) / [Repository](#)


Settings


 Profile


 Change password


 Security

 Linked accounts

 Applications

 Shared links


 **GitHub**

 **PhenixCollaboration/web**


DOI 10.5281/zenodo.3759876

GitHub / Releases


Create release ...

 v1.0 PhenixCollaboration/web: First release of the PHENIX DAP site

✓ Published

 DOI: 10.5281/zenodo.3759876

11 minutes ago

 First release of the PHENIX DAP site

Zenodo - GitHub panel - published release browser

The screenshot shows the Zenodo interface for a specific release. The header is blue with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. Below the header, the date 'April 21, 2020' is shown on the left, and 'Software' and 'Open Access' badges are on the right. The main title is 'PhenixCollaboration/web: First release of the PHENIX DAP site'. Below the title, the authors 'Maxim Potekhin; Ron Belmont; amolhj' are listed, followed by the description 'This is the first release of the PHENIX DAP website'. A 'Preview' section is expanded, showing a file tree for 'web-v1.0.zip'. The tree includes a root folder 'PhenixCollaboration-web-c9d991e' with sub-items: '.gitignore' (216 Bytes), 'Gemfile' (285 Bytes), 'LICENSE' (11.4 kB), 'README.md' (1.6 kB), '_about' folder, '_analysis' folder, '_config.yml' (1.3 kB), and '_data' folder. The '_data' folder contains 'acc' folder, 'vars.yml' (1.4 kB), 'detectors.yml' (2.6 kB), 'documents.yml' (3.5 kB), and 'gallery.yml' (4.5 kB). Other files in the '_about' folder include 'contact.md' (324 Bytes), 'dap.md' (1.3 kB), 'howto.md' (6.9 kB), and 'site.md' (2.5 kB). The '_analysis' folder contains 'overview.md' (114 Bytes).

zenodo Search Upload Communities

April 21, 2020 Software Open Access

PhenixCollaboration/web: First release of the PHENIX DAP site

Maxim Potekhin; Ron Belmont; amolhj

This is the first release of the PHENIX DAP website

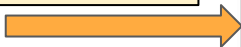
Preview

web-v1.0.zip

- PhenixCollaboration-web-c9d991e
 - .gitignore 216 Bytes
 - Gemfile 285 Bytes
 - LICENSE 11.4 kB
 - README.md 1.6 kB
 - _about
 - contact.md 324 Bytes
 - dap.md 1.3 kB
 - howto.md 6.9 kB
 - site.md 2.5 kB
 - _analysis
 - overview.md 114 Bytes
 - _config.yml 1.3 kB
 - _data
 - acc
 - vars.yml 1.4 kB
 - detectors.yml 2.6 kB
 - documents.yml 3.5 kB
 - gallery.yml 4.5 kB

Zenodo - an (elegant) way to reference software

Persistent, durable link to archived software, can be nicely embedded in any page.



The screenshot shows a Zenodo repository page. At the top, there's a list of recent changes: README.md (updated yesterday), _config.yml (8 days ago), and index.html (3 months ago). The main content is the README for 'The PHENIX Collaboration Data and Analysis Preservation Website'. It describes the site's purpose for long-term data and analysis preservation. A key feature is the release archival on Zenodo, with a DOI link: [DOI 10.5281/zenodo.3759876](https://doi.org/10.5281/zenodo.3759876). The page is divided into sections: SCOPE, TECHNICAL, For the Developer, and Gems. The SCOPE section mentions the use of the Jekyll static site generator and the site's role in preserving curated documentation. The TECHNICAL section includes a 'For the Developer' subsection with a list of dependencies for the Jekyll site.

README.md updated the README yesterday

_config.yml Reworked references to "assets", added flexlink to assets in the tab... 8 days ago

index.html Added includes, corrected README 3 months ago

README.md

The PHENIX Collaboration Data and Analysis Preservation Website

This website is designed to further the goals of the long-term Data and Analysis Preservation (DAP) of the PHENIX Collaboration.

The site is under heavy development and is not officially in production. At present, and until further notice, it is not expected to be 100% functional or have immediately useful content.

Release archival on Zenodo: [DOI 10.5281/zenodo.3759876](https://doi.org/10.5281/zenodo.3759876)

SCOPE

We use the Jekyll static site generator.

The site is intended to preserve curated documentation for the PHENIX experiment, including technical write-ups on the PHENIX software and its infrastructure. It is not a document server although it does host a limited number of documents (primarily in PDF formats) and as well as some diagrams.

Please note that the static nature of the site also implies lack of common database query functions at runtime, authentication and authorization etc. Where needed, such services will be hosted separately and links will be provided.

TECHNICAL

For the Developer

Please see the "how-to" section in the "About" menu of the site for the information being constantly updated.

Gems

Pay attention to the following dependencies (need to be installed and also included in the Gemfile in this folder):

```
gem "jekyll", "~> 4.0"
gem "jekyll-mentions", "~> 1.5", ">= 1.5.1"
gem "jekyll-sitemap", "~> 1.4"
gem "jekyll-redirect-from", "~> 0.16.0"
```

GitHub/Zenodo benefits for DAP

- Analysis note process: current policy as per the template is to create a tarball and indicate its location in HPSS
 - Very low-tech method of preserving the code, may be hard to find in a context other than the analysis note itself
 - Current solution - “leave for later”
- With GitHub (with optional Zenodo integration):
 - Analysis code/macros tagged with metadata, traceable and public (?)
 - Very easy to aggregate all information on one web page that works
 - Permanent/durable links in DAP documentation
- ...which brings us back to the policy issue

Policy

- Should we declare the analysis code public?
 - On the technical side, this would simplify things quite a bit
- If not, what are implications for DAP, especially for the DAP site which is public?
- I would argue that there is “security by obscurity” i.e. it would be impossible for an arbitrary person to make an arbitrary use of PHENIX code since it’s tied to the BNL infrastructure, and all access will need to be properly authorized